

Nemparaméteres statisztika

Györfi László

2013. június 12.

Tartalomjegyzék

1. Bevezetés	1
1.1. Miért becsülünk egy regressziós függvényt?	1
1.2. Hogyan becsülünk egy regressziós függvényt?	6
2. Partíciós beclés	11
2.1. Bevezetés	11
2.2. Stone tétele	12
2.3. Konzisztencia	15
2.4. Konvergenciasebesség	19
3. A magfüggvényes beclés	23
3.1. Bevezetés	23
3.2. Konzisztencia	24
3.3. A konvergencia sebessége	31
4. k legközelebbi szomszéd beclés	35
4.1. Bevezetés	35
4.2. Konzisztencia	37
4.3. A konvergencia-sebesség.	42
5. Idősorok predikciója	47
5.1. A predikciós probléma négyzetes hibával	47
5.2. Univerzálisan konzisztens partíciós stratégia	48
5.3. Univerzálisan konzisztens magfüggvényes stratégia	53
5.4. Univerzálisan konzisztens legközelebbi szomszéd stratégia	54
5.5. Univerzálisan konzisztens általánosított lineáris stratégia	55

6. Alakfelismerés	57
6.1. Bayes döntés	57
6.2. A Bayes döntés közelítése	61
6.3. Alakfelismerés idősorokra	63
7. Sűrűségfüggvénybecslés	69
7.1. Miért becsüljük sűrűségfüggvényt: az L_1 hiba	69
7.2. A hisztogram	72
7.3. Magfüggvényes sűrűségfüggvénybecslés	76
8. Egyszerű hipotézisek vizsgálata	79
8.1. α szintű tesztek	79
8.2. ϕ divergenciák	83
8.3. Ismételt megfigyelések	87
9. Hipotézisvizsgálat egyszerű null- és összetett alternatív hipotézis esetén	93
9.1. A variációs távolság és az I-divergencia	93
9.2. Az L_1 távolság nagy eltérése.	94
9.3. L_1 távolság alapú erősen konzisztens teszt	97
9.4. L_1 távolság alapú α szintű teszt	100
10. Homogenitás tesztelése	101
10.1. A tesztprobléma.	101
10.2. L_1 távolság alapú, erősen konzisztens teszt	101
10.3. L_1 távolság alapú α szintű teszt	105
11. Függetlenség tesztelése	107
11.1. A tesztprobléma	107
11.2. L_1 távolság alapú erősen konzisztens teszt	108
11.3. L_1 távolság alapú α szintű teszt	111
Bibliography	112

1. fejezet

Bevezetés

Ebben a fejezetben bevezetjük a regresszióbecslési problémát és a legfontosabb tulajdonságait, továbbá áttekintjük a különböző regresszióbecslési eljárásokat.

1.1. Miért becsülünk egy regressziós függvényt?

A regresszióanalízisben adott egy (\mathbf{X}, Y) véletlen vektor, ahol az \mathbf{X} megfigyelési vektor \mathbb{R}^d -ből veszi az értékeit, míg az Y valós. A regressziós problémában az érdekel bennünket, hogy mi a kapcsolat az Y (független) változó és az \mathbf{X} megfigyelési vektor között. Ez azt jelenti, hogy keressük azt a (mérhető) $f : \mathbb{R}^d \rightarrow \mathbb{R}$ függvényt, amelyre $f(\mathbf{X})$ egy „jó közelítése Y -nak,” azaz $f(\mathbf{X})$ legyen közel Y -hoz valamilyen értelemben, azaz $|f(\mathbf{X}) - Y|$ legyen „kicsi.” Mivel \mathbf{X} és Y véletlen, ezért $|f(\mathbf{X}) - Y|$ is véletlen, tehát egyáltalán nem világos, hogy mit értünk az alatt, hogy „kicsi $|f(\mathbf{X}) - Y|$ ”. A problémát úgy tesszük precízzé, hogy bevezetjük az f L_2 hibáját vagy négyzetes átlagos hibáját,

$$\mathbb{E}|f(\mathbf{X}) - Y|^2,$$

és azt kérjük, hogy ez legyen a lehető legkisebb.

Keressük azt az $m^* : \mathbb{R}^d \rightarrow \mathbb{R}$ (mérhető) függvényt, amelyre

$$\mathbb{E}|m^*(\mathbf{X}) - Y|^2 = \min_{f: \mathbb{R}^d \rightarrow \mathbb{R}} \mathbb{E}|f(\mathbf{X}) - Y|^2.$$

Megmutatjuk, hogy a keresett függvény az

$$m(\mathbf{x}) = \mathbb{E}\{Y | \mathbf{X} = \mathbf{x}\}$$

feltételes várható érték, amit *regressziós függvénynek* hívunk. Tetszőleges $f : \mathbb{R}^d \rightarrow \mathbb{R}$ függvényre

$$\begin{aligned}\mathbb{E}|f(\mathbf{X}) - Y|^2 &= \mathbb{E}|f(\mathbf{X}) - m(\mathbf{X}) + m(\mathbf{X}) - Y|^2 \\ &= \mathbb{E}|f(\mathbf{X}) - m(\mathbf{X})|^2 + \mathbb{E}|m(\mathbf{X}) - Y|^2,\end{aligned}$$

ahol kihasználtuk, hogy

$$\begin{aligned}\mathbb{E}\{(f(\mathbf{X}) - m(\mathbf{X}))(m(\mathbf{X}) - Y)\} \\ &= \mathbb{E}\{\mathbb{E}\{(f(\mathbf{X}) - m(\mathbf{X}))(m(\mathbf{X}) - Y)|\mathbf{X}\}\} \\ &= \mathbb{E}\{(f(\mathbf{X}) - m(\mathbf{X}))\mathbb{E}\{m(\mathbf{X}) - Y|\mathbf{X}\}\} \\ &= \mathbb{E}\{(f(\mathbf{X}) - m(\mathbf{X}))(m(\mathbf{X}) - m(\mathbf{X}))\} \\ &= 0.\end{aligned}$$

Tehát

$$\mathbb{E}|f(\mathbf{X}) - Y|^2 = \int_{\mathbb{R}^d} |f(\mathbf{x}) - m(\mathbf{x})|^2 \mu(d\mathbf{x}) + \mathbb{E}|m(\mathbf{X}) - Y|^2, \quad (1.1)$$

ahol μ jelöli az \mathbf{X} eloszlását. Az első tagot az f L_2 hibájának hívjuk. Ez mindig nemnegatív, és nulla akkor, ha $f(\mathbf{x}) = m(\mathbf{x})$. Következésképp $m^*(\mathbf{x}) = m(\mathbf{x})$, azaz az Y -nek L_2 hibában optimális közelítése valóban $m(\mathbf{X})$.

Egy gyakorlati alkalmazásban (\mathbf{X}, Y) együttes eloszlása és így maga a regressziós függvény ismeretlen, ezért nem tudjuk közelíteni Y -t az $m(\mathbf{X})$ segítségével. Ugyanakkor gyakran vannak adataink, amelyekből becsülhetjük a regressziós függvényt.

Jelölje (\mathbf{X}, Y) , (\mathbf{X}_1, Y_1) , $(\mathbf{X}_2, Y_2), \dots$ független, azonos eloszlású valószínűségi vektorváltozók egy sorozatát, ahol $\mathbb{E}Y^2 < \infty$. Legyen D_n az *adathalmaz*, amelyet az

$$D_n = \{(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)\}$$

definiál.

A regresszióbecslési problémában a D_n adatok felhasználásával szeretnénk becsülni az m regressziós függvényt. Az $m_n(\mathbf{x}) = m_n(\mathbf{x}, D_n)$ regresszióbecslés az \mathbf{x} és a D_n adatok függvénye.

A regresszióbecslés különbözik a regressziós függvénytől, tehát be kell vezetni egy hibakritériumot, amely összehasonlítja a regressziós függvényt és a regresszióbecslést. Az L_2 hiba lesz a természetes hibakritérium, mivel

$$\mathbb{E}\{|m_n(\mathbf{X}) - Y|^2 | D_n\} = \int_{\mathbb{R}^d} |m_n(\mathbf{x}) - m(\mathbf{x})|^2 \mu(d\mathbf{x}) + \mathbb{E}|m(\mathbf{X}) - Y|^2. \quad (1.2)$$

Tehát az m_n -nek az L_2 hibája akkor és csak akkor lesz kicsi, ha

$$\|m_n - m\|^2 = \int_{\mathbb{R}^d} |m_n(\mathbf{x}) - m(\mathbf{x})|^2 \mu(d\mathbf{x}) \quad (1.3)$$

közel nulla.

A regresszióbecslés klasszikus eljárása a paraméteres regresszióbecslés, amikor a regressziós függvény paraméteres szerkezete ismert, azaz csupán a regressziós függvény véges sok paraméterét kell becsülni. Példaként említhetjük a lineáris regressziót, amikor a regressziós függvény a komponensek lineáris függvénye $\mathbf{x} = (x^{(1)}, \dots, x^{(d)})^T$ -nek:

$$m(x^{(1)}, \dots, x^{(d)}) = a_0 + \sum_{i=1}^d a_i x^{(i)} \quad ((x^{(1)}, \dots, x^{(d)})^T \in \mathbb{R}^d)$$

valamilyen ismeretlen $a_0, \dots, a_d \in \mathbb{R}$ együtthatókkal (paraméterekkel). Ekkor az adatokból becsüljük a paramétereket, például úgy, hogy alkalmazzuk a legkisebb négyzetek elvét, ahol azt az a_0, \dots, a_d paramétervektort választjuk, amelyre az átlagos négyzetes hiba a legkisebb:

$$(\hat{a}_0, \dots, \hat{a}_d) = \arg \min_{a_0, \dots, a_d \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{j=1}^n \left| Y_j - a_0 - \sum_{i=1}^d a_i X_j^{(i)} \right|^2 \right\}.$$

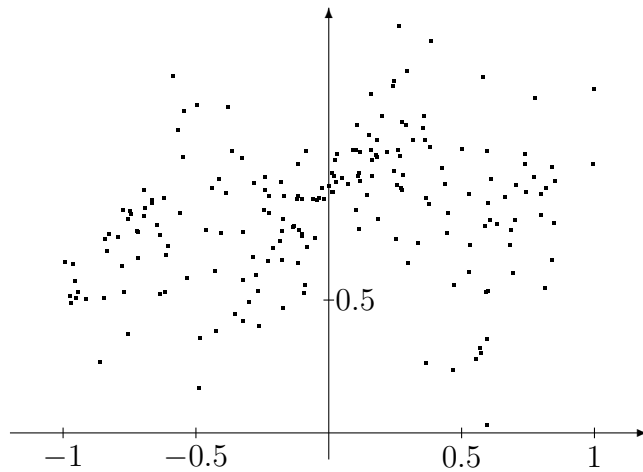
Itt $X_j^{(i)}$ jelöli az \mathbf{X}_j i -edik komponensét, és $\mathbf{z} = \arg \min_{\mathbf{x} \in D} f(\mathbf{x})$ és $\mathbf{z} \in D$ egy rövidítést úgy, hogy $f(\mathbf{z}) = \min_{\mathbf{x} \in D} f(\mathbf{x})$. Végül definiáljuk a paraméteres becslést:

$$\hat{m}_n(\mathbf{x}) = \hat{a}_0 + \sum_{i=1}^d \hat{a}_i x^{(i)}.$$

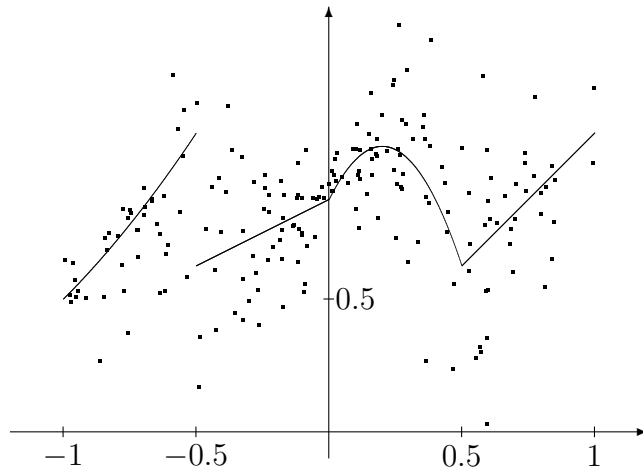
A paraméteres becslés csak néhány paramétertől függ, ezért már kis n mintanagyságnál is jó minőségű feltéve, hogy a paraméteres modell megfelelő. Azonban a paraméteres becslésnek van egy nagy hátránya, ugyanis a paraméteres becslés nem tudja jobban becsülni az ismeretlen regressziós függvényt, mint a legjobb függvényt a paraméteres függvényosztályban.

Egy-dimenziós $X = \mathbf{X}$ esetén most és a későbbiekben egy szimulált adathalmaznál illusztráljuk a különböző becsléseket. Ez a szimuláció $n = 200$ pontot tartalmaz úgy, hogy X standard normális eloszlású $[-1, 1]$ intervallumra megszorítva. A regressziós függvény szakaszonként polinomiális:

$$m(x) = \begin{cases} (x+2)^2/2 & \text{ha } -1 \leq x < -0.5, \\ x/2 + 0.875 & \text{ha } -0.5 \leq x < 0, \\ -5(x-0.2)^2 + 1.075 & \text{ha } 0 < x \leq 0.5, \\ x + 0.125 & \text{ha } 0.5 \leq x < 1. \end{cases}$$



1.1. ábra. Szimulált adatok.

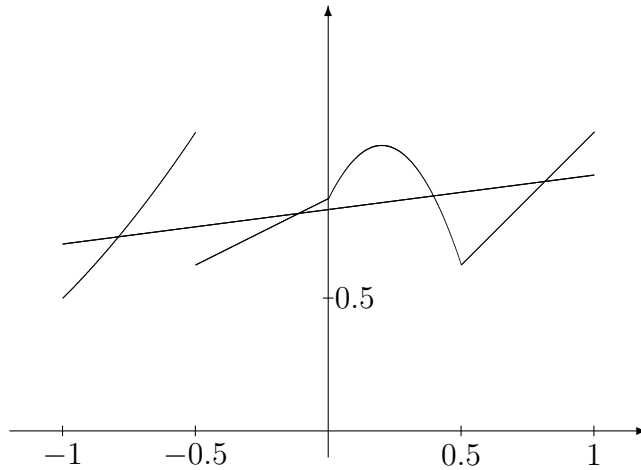


1.2. ábra. Adatpontok és a regressziós függvény.

Adott X esetén az $Y - m(X)$ feltételes eloszlása nulla várható értékű normális a következő szórással:

$$\sigma(X) = 0.2 - 0.1 \cos(2\pi X).$$

Az 1.1 ábra mutatja a mintapontokat. Ebben a példában szemmel nem látható, hogy az adatpontok mögött mi a regressziós függvény. Az 1.2 ábra mutatja együtt az adatokat



1.3. ábra. Lineáris regressziós becslés.

és a regressziós függvényt. Az 1.3 ábrában a lineáris becslés és a regressziós függvény látható a szimulált adatok esetén.

Többváltozós \mathbf{X} esetén nehezen lehet vizualizálni az adatokat, és ekkor a jó minőségű paraméteres modellt is bonyolult felépíteni, és egy rossz modelltől levezetett regressziós becslés általában gyenge minőségű. A paraméteres modellezésnek ez a rugalmatlansága kerülhető el, ha nemparaméteres módszereket alkalmazunk, amikor nem tételezünk fel semmit az ismeretlen regressziós függvényről.

A későbbiekben két konvergenciatípust vizsgálunk. Az első és egyben gyengébb konzisztencia esetén azt követeljük, hogy az L_2 hiba

$$\int |m_n(\mathbf{x}) - m(\mathbf{x})|^2 \mu(d\mathbf{x}).$$

várható értéke tartson nullához.

1.1 definíció Az $\{m_n\}$ regressziós becslést **gyengén konzisztensnek** nevezzük, ha

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \int (m_n(\mathbf{x}) - m(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} = 0$$

az (\mathbf{X}, Y) bizonyos eloszlásaira.

1.2 definíció Az $\{m_n\}$ regressziós becslést **erősen konzisztensnek** nevezzük, ha

$$\lim_{n \rightarrow \infty} \int (m_n(\mathbf{x}) - m(\mathbf{x}))^2 \mu(d\mathbf{x}) = 0 \quad 1 \text{ valószínűséggel}$$

az (\mathbf{X}, Y) bizonyos eloszlásaira.

Lehetséges, hogy egy regressziós becslés konzisztens bizonyos eloszlásokra, és másokra nem. A következő fejezetekben olyan m_n regressziós becsléseket vizsgálunk, amelyek univerzálisan konzisztensek, ami azt jelenti, hogy konzisztensek minden olyan esetben, amikor a regressziós probléma értelmezett.

1.3 definíció Az $\{m_n\}$ regressziós becslést **gyengén univerzálisan konzisztensnek** nevezzük, ha gyengén konzisztens az (\mathbf{X}, Y) minden olyan eloszlására, amikor $\mathbb{E}\{Y^2\} < \infty$.

1.4 definíció Az $\{m_n\}$ regressziós becslést **erősen univerzálisan konzisztensnek** nevezzük, ha erősen konzisztens az (\mathbf{X}, Y) minden olyan eloszlására, amikor $\mathbb{E}\{Y^2\} < \infty$.

Ha egy becslés univerzálisan konzisztens, akkor az L_2 hiba nullához tart függetlenül az (\mathbf{X}, Y) igazi eloszlásától. Ugyanakkor ez semmit nem mond a konvergencia sebességéről.

A későbbiekben az L_2 hiba várható értékének

$$\mathbb{E} \int |m_n(\mathbf{x}) - m(\mathbf{x})|^2 \mu(d\mathbf{x}). \quad (1.4)$$

a konvergencia-sebességét is vizsgáljuk.

Természetesen adódik az a kérdés, hogy hogyan konstruálhatunk olyan becslést, amelyre (1.4) nullához tart egy nemtriviális sebességgel az (\mathbf{X}, Y) minden eloszlására. Sajnos ilyen becslés nem létezik, mivel tetszőleges becslés esetén a konvergencia sebessége akármilyen lassú lehet. A nemtriviális konvergencia-sebességhez korlátozni kell az (\mathbf{X}, Y) eloszlását, például fel kell tenni a regressziós függvény folytonosságát.

1.2. Hogyan becsüljünk egy regressziós függvényt?

Ebben a szakaszban nemparaméteres regresszióbecslés két alapvető elvét mutatjuk meg: **lokális átlagolás és empirikus hibaminimalizálás.**

Emlékeztetünk arra, hogy a regressziós függvényt a

$$m(\mathbf{x}) = \mathbb{E}\{Y \mid \mathbf{X} = \mathbf{x}\}$$

feltételes várható értékkel definiáltuk.

Ha \mathbf{x} az \mathbf{X} egy atomja, azaz $\mathbb{P}\{\mathbf{X} = \mathbf{x}\} > 0$, akkor a feltételes várható értéket hagyományosan a következő módon definiáljuk:

$$\mathbb{E}\{Y \mid \mathbf{X} = \mathbf{x}\} = \frac{\mathbb{E}\{Y \mathbb{I}_{\{\mathbf{X}=\mathbf{x}\}}\}}{\mathbb{P}\{\mathbf{X} = \mathbf{x}\}},$$

ahol \mathbb{I}_A az A halmaz indikátorát jelöli. Ebben a definícióban a számlálót az

$$\frac{1}{n} \sum_{i=1}^n Y_i \mathbb{I}_{\{\mathbf{X}_i=\mathbf{x}\}}$$

átlaggal becsülhetjük, míg a nevezőt az

$$\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i=\mathbf{x}\}}$$

relatív gyakorisággal, így a természetesen adódó regresszióbecslés az

$$m_n(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{\{\mathbf{X}_i=\mathbf{x}\}}}{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i=\mathbf{x}\}}}.$$

Általános esetben előfordulhat, hogy $\mathbb{P}\{\mathbf{X} = \mathbf{x}\} = 0$. Ekkor egyrészt mértékelméleti technikákkal bevezethető a feltételes várható érték (lásd Appendix a Devroye, Györfi, and Lugosi [1996] könyvben). Sajnos ez a definíció nehézkesen használható a statisztikában. Egy ekvivalens definícióra juthatunk a

$$\mathbb{E}\{Y \mid \mathbf{X} = \mathbf{x}\} = \lim_{h \rightarrow 0} \frac{\mathbb{E}\{Y \mathbb{I}_{\{\|\mathbf{X}-\mathbf{x}\| \leq h\}}\}}{\mathbb{P}\{\|\mathbf{X} - \mathbf{x}\| \leq h\}}$$

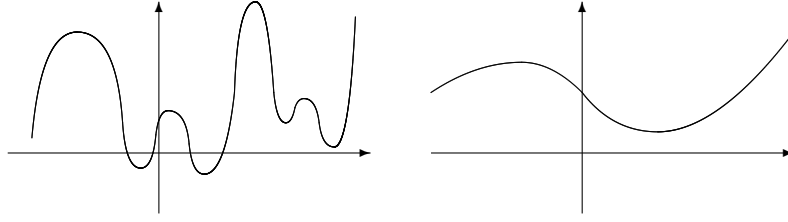
tulajdonság alapján, amelyből a következő becslés vezethető le:

$$m_n(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{\{\|\mathbf{X}_i-\mathbf{x}\| \leq h\}}}{\sum_{i=1}^n \mathbb{I}_{\{\|\mathbf{X}_i-\mathbf{x}\| \leq h\}}}.$$

Ezt a becslést naiv magfüggvényes becslésnek hívjuk, amelyik egy lokális átlagolás elvén működő becslés.

A *lokális átlagolásos becslések* általános alakja az

$$m_n(\mathbf{x}) = \sum_{i=1}^n W_{n,i}(\mathbf{x}) \cdot Y_i,$$



1.4. ábra. A jobboldali becslés ésszerűbbnek tűnik, mint a baloldali, amely csupán interpolál.

ahol a $W_{n,i}(\mathbf{x}) = W_{n,i}(\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_n) \in \mathbb{R}$ súlyok az $\mathbf{X}_1, \dots, \mathbf{X}_n$ függvényei. Általában a $W_{n,i}(\mathbf{x})$ súlyok nemnativak, és „kicsik”, ha \mathbf{X}_i „távol” van \mathbf{x} -től. Lokális átlagolásos becslésekre példa a *partíciós becslés*, a *magfüggvényes becslés* és a *k legközelebbi szomszéd becslés*. Ezeknek a becsléseknek az alapvető tulajdonságait a következő fejezetekben mutatjuk be.

Az *empirikus hibaminimalizálás*on alapuló becslési módszereknél adott \mathbb{R}^d -n értelmezett függvényeknek egy \mathcal{F}_n családja. Ekkor a becslést a következő módon definiáljuk:

$$m_n(\cdot) = \arg \min_{f \in \mathcal{F}_n} \left\{ \frac{1}{n} \sum_{i=1}^n |f(\mathbf{X}_i) - Y_i|^2 \right\}. \quad (1.5)$$

Az empirikus hibaminimalizálásban alapuló becslés minimalizálja az

$$\frac{1}{n} \sum_{i=1}^n |f(\mathbf{X}_i) - Y_i|^2 \quad (1.6)$$

empirikus L_2 hibát \mathcal{F}_n -en. Vegyük észre, hogy nincs értelme annak, hogy (1.6)-t minimalizáljunk minden (mérhető) f függvényre, mivel a kapott függvény csak interpolálja az adatokat, és nem egy ésszerű becslés. Ezért meg kell szorítani az \mathcal{F}_n családot. Az \mathcal{F}_n családra egy fontos példa az általánosított lineáris becslés. Legyenek $\{\phi_j\}_{j=1}^{\infty}$ valós értékű, \mathbb{R}^d -én értelmezett függvények, és legyen

$$\mathcal{F}_n = \left\{ f; f = \sum_{j=1}^{\ell_n} c_j \phi_j \right\}.$$

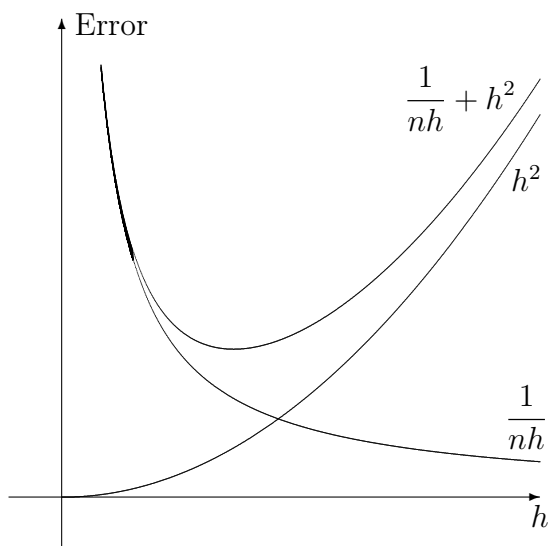
Akkor az általánosított lineáris becslést a következő módon definiáljuk:

$$\begin{aligned} m_n(\cdot) &= \arg \min_{f \in \mathcal{F}_n} \left\{ \frac{1}{n} \sum_{i=1}^n (f(\mathbf{X}_i) - Y_i)^2 \right\} \\ &= \arg \min_{c_1, \dots, c_{\ell_n}} \left\{ \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^{\ell_n} c_j \phi_j(\mathbf{X}_i) - Y_i \right)^2 \right\}. \end{aligned}$$

Legyen m_n egy tetszőleges becslés, akkor egy $\mathbf{x} \in \mathbb{R}^d$ pontban az átlagos négyzetes hibát felírhatjuk

$$\begin{aligned} &\mathbb{E}\{|m_n(\mathbf{x}) - m(\mathbf{x})|^2\} \\ &= \mathbb{E}\{|m_n(\mathbf{x}) - \mathbb{E}\{m_n(\mathbf{x})\}|^2\} + |\mathbb{E}\{m_n(\mathbf{x})\} - m(\mathbf{x})|^2 \\ &= \text{Var}(m_n(\mathbf{x})) + |\text{bias}(m_n(\mathbf{x}))|^2, \end{aligned}$$

formában, ahol $\text{Var}(m_n(\mathbf{x}))$ az $m_n(\mathbf{x})$ valószínűségi változó szórásnégyzete, és $\text{bias}(m_n(\mathbf{x}))$



1.5. ábra. A szórás és a torzítás kapcsolata.

a torzítása. Ebből következik egy felbontás az L_2 hiba várható értékére:

$$\begin{aligned} & \mathbb{E} \left\{ \int |m_n(\mathbf{x}) - m(\mathbf{x})|^2 \mu(d\mathbf{x}) \right\} \\ &= \int \mathbb{E}\{|m_n(\mathbf{x}) - m(\mathbf{x})|^2\} \mu(d\mathbf{x}) \\ &= \int \text{Var}(m_n(\mathbf{x})) \mu(d\mathbf{x}) + \int |\text{bias}(m_n(\mathbf{x}))|^2 \mu(d\mathbf{x}). \end{aligned}$$

Naiv magfüggvényes becslés esetén az 1.5 ábra illusztrálja a szórásnégyzet és a torzításnégyzet függését h -tól, ugyanis bizonyos regularitási feltételek esetén

$$\int_{\mathbb{R}^d} \text{Var}(m_n(\mathbf{x})) \mu(d\mathbf{x}) = c_1 \frac{1}{nh^d} + o\left(\frac{1}{nh^d}\right)$$

és

$$\int_{\mathbb{R}^d} |\text{bias}(m_n(\mathbf{x}))|^2 \mu(d\mathbf{x}) = c_2 h^2 + o(h^2).$$

2. fejezet

Partíciós becslés

2.1. Bevezetés

A következő fejezetekben áttekintjük a legfontosabb lokális átlagolásos becsléseket. További részletek megtalálhatók a Györfi *et al.* [2002] könyvben.

Legyen $\mathcal{P}_n = \{A_{n,1}, A_{n,2}, \dots\}$ az \mathbb{R}^d egy partíciója, és tetszőleges $\mathbf{x} \in \mathbb{R}^d$ esetén $A_n(\mathbf{x})$ jelölje a \mathcal{P}_n -nek azt az $A_{n,j}$ celláját, amibe \mathbf{x} esik. A partíciós regresszióbecslést a

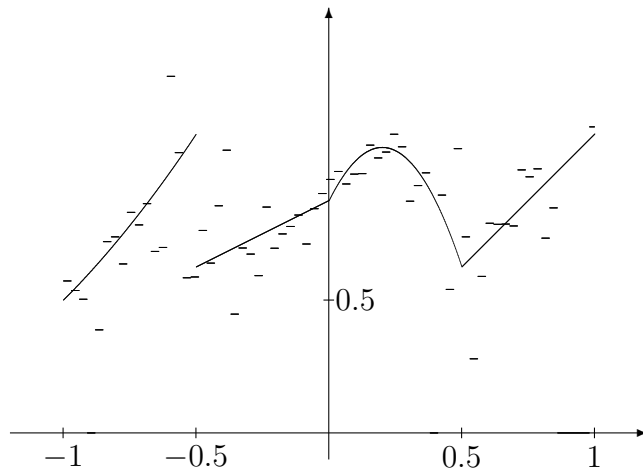
$$m_n(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{x})\}}}{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{x})\}}}$$

hányadossal definiáljuk úgy, hogy definíció szerint $0/0 = 0$. Ez azt jelenti, hogy a partíciós becslés lokális átlagolás típusú, azaz adott \mathbf{x} esetén átlagoljuk azokat az Y_i -ket, amelyekre \mathbf{X}_i ugyanabba a cellába esik, mint amibe \mathbf{x} esik.

A becslés legegyszerűbb speciális esetében $d = 1$ és az $A_{n,j}$ cellák $h = h_n$ hosszúságú intervallumok. A 2.1 – 2.3 ábrák a becslést illusztrálják különböző h esetén az 1. fejezetben leírt szimulált adatokon. Az első ábrán h túl kicsi, mert nagy a szórás, a másodikban a h választása lényegében jó, míg a harmadikban túl nagy, mivel a torzítás nagy.

$d > 1$ estén a partíció állhat h_n oldalhosszúságú kockákból, azaz az $A_{n,j}$ cellák h_n^d térfogatú kockák, vagy az $A_{n,j}$ cellák lehetnek téglák h_{n1}, \dots, h_{nd} oldalhosszúságokkal. Illusztrációképpen két dimenziós, korrelált normális eloszlás esetén a 2.4 ábrán a partíció kockás, míg 2.5 ábra esetén téglás.

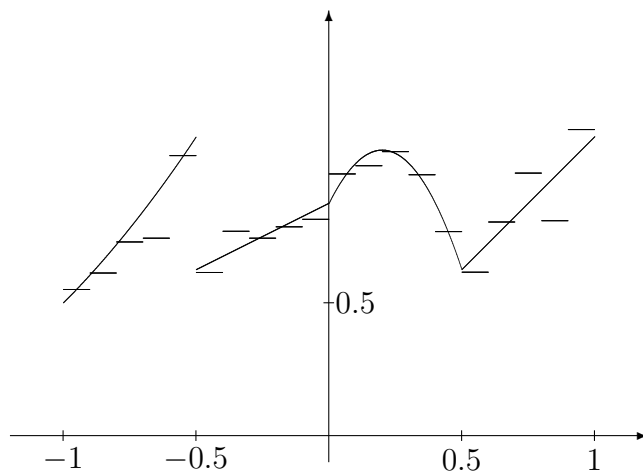
A partíció lehet adatfüggő. A 2.6 ábra mutat egy ilyen partíciót, ahol minden cella ugyanannyi pontot tartalmaz. Az ilyen partíciót statisztikailag ekvivalens blokkoknak hívjuk.



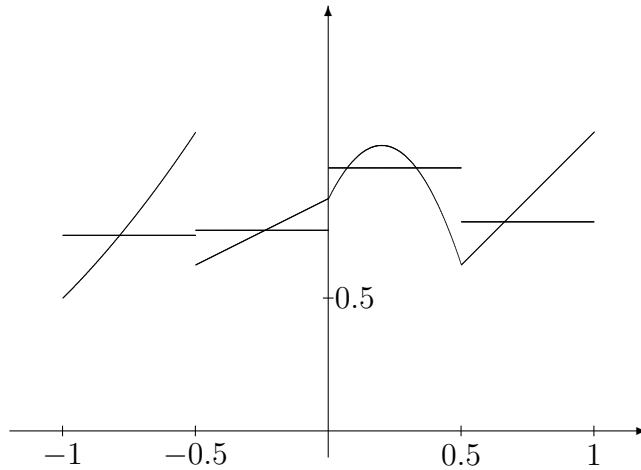
2.1. ábra. Nagy a szórás: $h = 0.03$, L_2 hiba = 0.062433.

2.2. Stone tétele

Ebben a szakaszban a partíciós becslés gyenge univerzális konzisztenciáját mutatjuk meg. A bizonyítás Stone tételére hivatkozik (2.2.1. tétel), ami egy igen hatékony és általános eszköz lokális átlagolás típusú becslések analizisében.



2.2. ábra. Jó választás: $h = 0.1$, L_2 hiba = 0.003642.



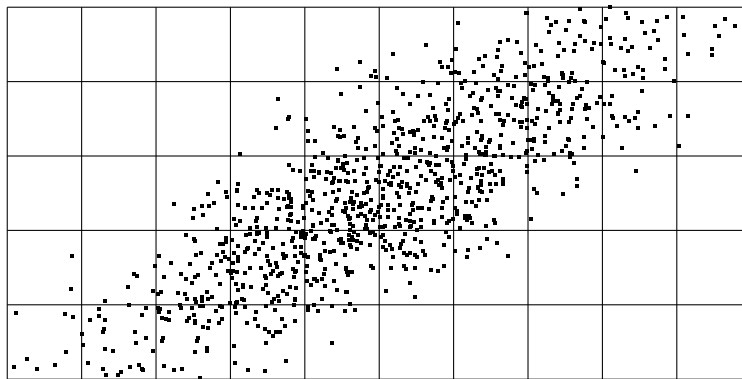
2.3. ábra. Nagy a torzítás: $h = 0.5$, L_2 hiba = 0.013208.

A lokális átlagolás típusú regresszióbecslések

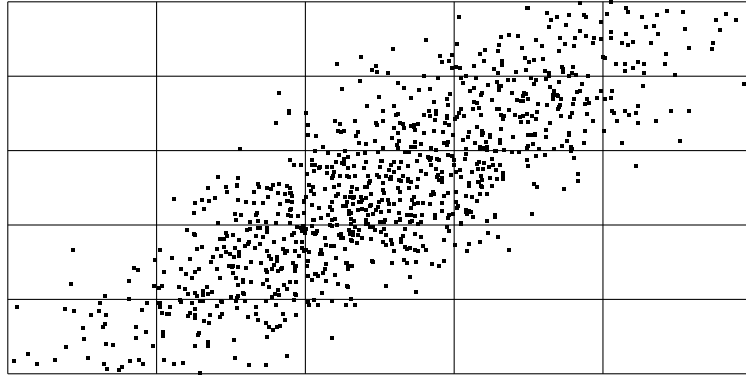
$$m_n(\mathbf{x}) = \sum_{i=1}^n W_{ni}(\mathbf{x}) \cdot Y_i,$$

alakúak, ahol a $W_{n,i}(\mathbf{x}) = W_{n,i}(\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_n) \in \mathbb{R}$ súlyok függenek $\mathbf{X}_1, \dots, \mathbf{X}_n$ -től.

Általában a $W_{n,i}(\mathbf{x})$ súlyok nemnegatívak és „kicsik”, ha \mathbf{X}_i „messze” van \mathbf{x} -től. A következő tétel megadja azokat az általános feltételeket, amelyek garantálják a gyenge a lokális átlagolás típusú regresszióbecslések univerzális konzisztenciáját.



2.4. ábra. Kockás partíció.

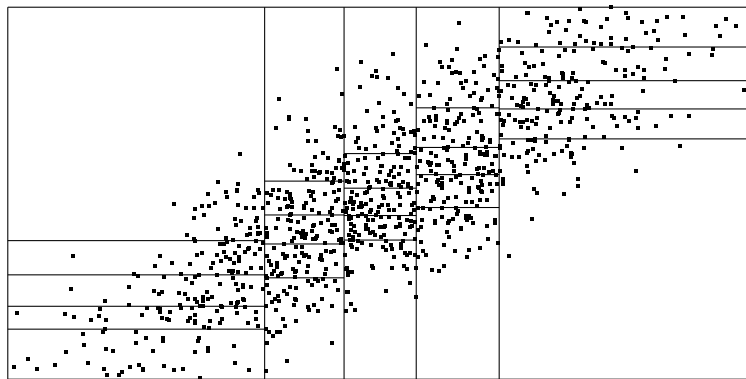


2.5. ábra. Téglás partíció.

2.2.1. tétel (STONE TÉTEL). *Tegyük fel, hogy az \mathbf{X} tetszőleges eloszlására teljesülnek az alábbi feltételek:*

- (i) *Létezik egy c konstans úgy, hogy minden f nemnegatív értékű, mérhető függvényre, amelyre $\mathbb{E}f(\mathbf{X}) < \infty$, arra minden n -re,*

$$\mathbb{E} \left\{ \sum_{i=1}^n |W_{n,i}(\mathbf{X})| f(\mathbf{X}_i) \right\} \leq c \mathbb{E}f(\mathbf{X}).$$



2.6. ábra. Statisztikailag ekvivalens blokkok.

(ii) Létezik egy $D \geq 1$ konstans, amelyre

$$\mathbb{P} \left\{ \sum_{i=1}^n |W_{n,i}(\mathbf{X})| \leq D \right\} = 1,$$

minden n -re.

(iii) Minden $a > 0$ -ra,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \sum_{i=1}^n |W_{n,i}(\mathbf{X})| \mathbb{I}_{\{\|\mathbf{x}_i - \mathbf{x}\| > a\}} \right\} = 0.$$

(iv)

$$\sum_{i=1}^n W_{n,i}(\mathbf{X}) \rightarrow 1$$

(v) sztochasztikusan.

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \sum_{i=1}^n W_{n,i}(\mathbf{X})^2 \right\} = 0.$$

Akkor a szóbanforgó m_n regresszióbecslés gyengén univerzálisan konzisztens, azaz

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \int (m_n(\mathbf{x}) - m(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} = 0$$

az (\mathbf{X}, Y) minden olyan eloszlására, ahol $\mathbb{E}Y^2 < \infty$.

A bizonyítás megtalálható a Györfi *et al.* [2002] könyvben.

2.3. Konzisztencia

Ebben a szakaszban megmutatjuk a partíciós becslés gyenge univerzális konzisztenciáját.

2.3.2. tétel *Tegyük fel, hogy tetszőleges origó középpű S gömbre*

$$\lim_{n \rightarrow \infty} \max_{j: A_{n,j} \cap S \neq \emptyset} \text{diam}(A_{n,j}) = 0 \tag{2.1}$$

és

$$\lim_{n \rightarrow \infty} \frac{|\{j : A_{n,j} \cap S \neq \emptyset\}|}{n} = 0. \tag{2.2}$$

Akkor a partíciós becslés gyengén univerzálisan konzisztens.

Kockás partíció esetén

$$\lim_{n \rightarrow \infty} h_n = 0 \text{ és } \lim_{n \rightarrow \infty} nh_n^d = \infty$$

ekvivalensek (2.1)-gyel és (2.2)-vel.

A 2.3.2. tételt úgy bizonyítjuk, hogy ellenőrizzük Stone tétel feltételeit. Ehhez szükség van egy technikai lemmára. Legyen $B(n, p)$ egy binomiális eloszlású valószínűségi változó n és $0 \leq p \leq 1$ paraméterekkel, azaz

$$\mathbb{P}\{B(n, p) = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

2.1 lemma *Legyen $B(n, p)$ egy binomiális eloszlású valószínűségi változó n és $0 \leq p \leq 1$ paraméterekkel. Akkor*

(i)

$$\mathbb{E} \left\{ \frac{1}{1 + B(n, p)} \right\} \leq \frac{1}{(n+1)p},$$

(ii)

$$\mathbb{E} \left\{ \frac{1}{B(n, p)} \mathbb{I}_{\{B(n, p) > 0\}} \right\} \leq \frac{2}{(n+1)p}.$$

BIZONYÍTÁS. (i) következik az alábbi egyszerű számolásból:

$$\begin{aligned} \mathbb{E} \left\{ \frac{1}{1 + B(n, p)} \right\} &= \sum_{k=0}^n \frac{1}{k+1} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{1}{(n+1)p} \sum_{k=0}^n \binom{n+1}{k+1} p^{k+1} (1-p)^{n-k} \\ &\leq \frac{1}{(n+1)p} \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n-k+1} \\ &= \frac{1}{(n+1)p} (p + (1-p))^{n+1} \\ &= \frac{1}{(n+1)p}. \end{aligned}$$

(ii) esetén

$$\mathbb{E} \left\{ \frac{1}{B(n,p)} \mathbb{I}_{\{B(n,p)>0\}} \right\} \leq \mathbb{E} \left\{ \frac{2}{1+B(n,p)} \right\} \leq \frac{2}{(n+1)p}$$

az (i) miatt. □

A 2.3.2. TÉTEL BIZONYÍTÁSA. A bizonyítás abból áll, hogy ellenőrizzük a Stone tétel feltételeit (2.2.1. tétel). Mivel definíció szerint $0/0 = 0$, ezért

$$W_{n,i}(\mathbf{x}) = \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{x})\}} / \sum_{l=1}^n \mathbb{I}_{\{\mathbf{X}_l \in A_n(\mathbf{x})\}}.$$

(i) ellenőrzéséhez elég megmutatni, hogy létezik egy $c > 0$ konstans úgy, hogy minden f nemnegatív függvényre, amelyre $\mathbb{E}f(\mathbf{X}) < \infty$, teljesül

$$\mathbb{E} \left\{ \sum_{i=1}^n f(\mathbf{X}_i) \frac{\mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{X})\}}}{\sum_{l=1}^n \mathbb{I}_{\{\mathbf{X}_l \in A_n(\mathbf{X})\}}} \right\} \leq c \mathbb{E}f(\mathbf{X}).$$

Vegyük észre, hogy

$$\begin{aligned} & \mathbb{E} \left\{ \sum_{i=1}^n f(\mathbf{X}_i) \frac{\mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{X})\}}}{\sum_{l=1}^n \mathbb{I}_{\{\mathbf{X}_l \in A_n(\mathbf{X})\}}} \right\} \\ &= \sum_{i=1}^n \mathbb{E} \left\{ f(\mathbf{X}_i) \frac{\mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{X})\}}}{1 + \sum_{l \neq i} \mathbb{I}_{\{\mathbf{X}_l \in A_n(\mathbf{X})\}}} \right\} \\ &= n \mathbb{E} \left\{ f(\mathbf{X}_1) \mathbb{I}_{\{\mathbf{X}_1 \in A_n(\mathbf{X})\}} \frac{1}{1 + \sum_{l \neq 1} \mathbb{I}_{\{\mathbf{X}_l \in A_n(\mathbf{X})\}}} \right\} \\ &= n \mathbb{E} \left\{ \mathbb{E} \left\{ f(\mathbf{X}_1) \mathbb{I}_{\{\mathbf{X}_1 \in A_n(\mathbf{X})\}} \frac{1}{1 + \sum_{l=2}^n \mathbb{I}_{\{\mathbf{X}_l \in A_n(\mathbf{X})\}}} \middle| \mathbf{X}, \mathbf{X}_1 \right\} \right\} \\ &= n \mathbb{E} \left\{ f(\mathbf{X}_1) \mathbb{I}_{\{\mathbf{X}_1 \in A_n(\mathbf{X})\}} \mathbb{E} \left\{ \frac{1}{1 + \sum_{l=2}^n \mathbb{I}_{\{\mathbf{X}_l \in A_n(\mathbf{X})\}}} \middle| \mathbf{X}, \mathbf{X}_1 \right\} \right\} \\ &= n \mathbb{E} \left\{ f(\mathbf{X}_1) \mathbb{I}_{\{\mathbf{X}_1 \in A_n(\mathbf{X})\}} \mathbb{E} \left\{ \frac{1}{1 + \sum_{l=2}^n \mathbb{I}_{\{\mathbf{X}_l \in A_n(\mathbf{X})\}}} \middle| \mathbf{X} \right\} \right\}, \end{aligned}$$

ahol kihasználtuk, hogy $\mathbf{X}, \mathbf{X}_1, \dots, \mathbf{X}_n$ függetlenek. A 2.1 lemma miatt a fenti várható érték felülről korlátozható:

$$\begin{aligned} n\mathbb{E} \left\{ f(\mathbf{X}_1) \mathbb{I}_{\{\mathbf{X}_1 \in A_n(\mathbf{X})\}} \frac{1}{n\mu(A_n(\mathbf{X}))} \right\} &= \sum_j \mathbb{P}\{\mathbf{X} \in A_{nj}\} \int_{A_{nj}} f(u) \mu(du) \frac{1}{\mu(A_{nj})} \\ &= \int_{\mathbb{R}^d} f(u) \mu(du) = \mathbb{E}f(\mathbf{X}), \end{aligned}$$

tehát az (i) feltétel teljesül $c = 1$ konstanssal. Könnyen látható, hogy az (ii) feltétel teljesül $D = 1$ konstanssal. Az (iii) feltétel ellenőrzéséhez válasszunk egy origó közepű S gömböt. Akkor a (2.1) feltétel miatt elegendően nagy n -re és $A_{n,j} \cap S \neq \emptyset$ -re $\text{diam}(A_{n,j}) < a$. Ekkor $\mathbf{X} \in S$ és $\|\mathbf{X}_i - \mathbf{X}\| > a$ esetén $\mathbf{X}_i \notin A_n(\mathbf{X})$, ezért

$$\begin{aligned} \mathbb{I}_{\{\mathbf{X} \in S\}} \sum_{i=1}^n W_{n,i}(\mathbf{X}) \mathbb{I}_{\{\|\mathbf{X}_i - \mathbf{X}\| > a\}} &= \mathbb{I}_{\{\mathbf{X} \in S\}} \frac{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{X}), \|\mathbf{X} - \mathbf{X}_i\| > a\}}}{n\mu_n(A_n(\mathbf{X}))} \\ &= \mathbb{I}_{\{\mathbf{X} \in S\}} \frac{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{X}), \mathbf{X}_i \notin A_n(\mathbf{X}), \|\mathbf{X} - \mathbf{X}_i\| > a\}}}{n\mu_n(A_n(\mathbf{X}))} \\ &= 0. \end{aligned}$$

Ebből következik, hogy

$$\limsup_n \mathbb{E} \sum_{i=1}^n W_{n,i}(\mathbf{X}) \mathbb{I}_{\{\|\mathbf{X}_i - \mathbf{X}\| > a\}} \leq \mu(S^c).$$

A (iv) feltételhez vegyük észre, hogy

$$\begin{aligned} \mathbb{P} \left\{ \sum_{i=1}^n W_{n,i}(\mathbf{X}) \neq 1 \right\} &= \mathbb{P} \{ \mu_n(A_n(\mathbf{X})) = 0 \} \\ &= \sum_j \mathbb{P} \{ \mathbf{X} \in A_{n,j}, \mu_n(A_{n,j}) = 0 \} \\ &= \sum_j \mu(A_{n,j}) (1 - \mu(A_{n,j}))^n \\ &\leq \sum_{j: A_{n,j} \cap S = \emptyset} \mu(A_{n,j}) + \sum_{j: A_{n,j} \cap S \neq \emptyset} \mu(A_{n,j}) (1 - \mu(A_{n,j}))^n. \end{aligned}$$

Az elemi

$$x(1-x)^n \leq xe^{-nx} \leq \frac{1}{en} \quad (0 \leq x \leq 1)$$

egyenlőtlenségből következik, hogy

$$\mathbb{P} \left\{ \sum_{i=1}^n W_{n,i}(\mathbf{X}) \neq 1 \right\} \leq \mu(S^c) + \frac{1}{en} |\{j : A_{n,j} \cap S \neq \emptyset\}|.$$

A jobboldal első tagja tetszőlegesen kicsi lehet az S gömb választásával, míg a második tag nullához tart a (2.2) feltétel miatt. A (v) feltétel igazolásához megjegyezzük, hogy

$$\sum_{i=1}^n W_{n,i}(\mathbf{x})^2 = \begin{cases} \frac{1}{\sum_{l=1}^n \mathbb{I}_{\{\mathbf{x}_l \in A_n(\mathbf{x})\}}} & \text{if } \mu_n(A_n(\mathbf{x})) > 0, \\ 0 & \text{if } \mu_n(A_n(\mathbf{x})) = 0, \end{cases}$$

amiből következik, hogy

$$\begin{aligned} & \mathbb{E} \left\{ \sum_{i=1}^n W_{n,i}(\mathbf{X})^2 \right\} \\ & \leq \mathbb{P}\{\mathbf{X} \in S^c\} + \sum_{j:A_{n,j} \cap S \neq \emptyset} \mathbb{E} \left\{ \mathbb{I}_{\{\mathbf{x} \in A_{n,j}\}} \frac{1}{n\mu_n(A_{n,j})} \mathbb{I}_{\{\mu_n(A_{n,j}) > 0\}} \right\} \\ & \leq \mu(S^c) + \sum_{j:A_{n,j} \cap S \neq \emptyset} \mu(A_{n,j}) \frac{2}{n\mu(A_{n,j})} \\ & \hspace{15em} (2.1 \text{ lemma miatt}) \\ & = \mu(S^c) + \frac{2}{n} |\{j : A_{n,j} \cap S \neq \emptyset\}|. \end{aligned}$$

Az előzőhöz hasonló gondolatmenettel a bizonyítás teljes. □

2.4. Konvergenciasebesség

Ebben a szakaszban a $\mathbb{E}\|m_n - m\|^2$ konvergencia-sebességét számoljuk ki kockás partíció és Lipschitz folytonos regressziófüggvény esetén.

2.4.3. tétel h_n oldalhosszúságú kockás partíció esetén tegyük fel, hogy

$$\begin{aligned} \text{Var}(Y|\mathbf{X} = \mathbf{x}) & \leq \sigma^2, \quad \mathbf{x} \in \mathbb{R}^d, \\ |m(\mathbf{x}) - m(\mathbf{z})| & \leq C\|\mathbf{x} - \mathbf{z}\|, \quad \mathbf{x}, \mathbf{z} \in \mathbb{R}^d, \end{aligned} \tag{2.3}$$

és \mathbf{X} tartója a kompakt S halmaz. Akkor

$$\mathbb{E}\|m_n - m\|^2 \leq \hat{c} \frac{\sigma^2 + \sup_{z \in S} |m(z)|^2}{n \cdot h_n^d} + d \cdot C^2 \cdot h_n^2,$$

ahol \hat{c} csak d -től és az S átmérőjétől függ, tehát

$$h_n = c' \left(\frac{\sigma^2 + \sup_{z \in S} |m(z)|^2}{C^2} \right)^{1/(d+2)} n^{-\frac{1}{d+2}}$$

választás esetén

$$\mathbb{E}\|m_n - m\|^2 \leq c'' \left(\sigma^2 + \sup_{z \in S} |m(z)|^2 \right)^{2/(d+2)} C^{2d/(d+2)} n^{-2/(d+2)}.$$

BIZONYÍTÁS. Legyen

$$\hat{m}_n(\mathbf{x}) = \mathbb{E}\{m_n(\mathbf{x}) | \mathbf{X}_1, \dots, \mathbf{X}_n\} = \frac{\sum_{i=1}^n m(\mathbf{X}_i) \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{x})\}}}{n\mu_n(A_n(\mathbf{x}))}.$$

Akkor

$$\begin{aligned} & \mathbb{E}\{(m_n(\mathbf{x}) - m(\mathbf{x}))^2 | \mathbf{X}_1, \dots, \mathbf{X}_n\} \\ &= \mathbb{E}\{(m_n(\mathbf{x}) - \hat{m}_n(\mathbf{x}))^2 | \mathbf{X}_1, \dots, \mathbf{X}_n\} + (\hat{m}_n(\mathbf{x}) - m(\mathbf{x}))^2. \end{aligned} \quad (2.4)$$

Egyrészt a szórás típusú tagra azt kapjuk, hogy

$$\begin{aligned} & \mathbb{E}\{(m_n(\mathbf{x}) - \hat{m}_n(\mathbf{x}))^2 | \mathbf{X}_1, \dots, \mathbf{X}_n\} \\ &= \mathbb{E} \left\{ \left(\frac{\sum_{i=1}^n (Y_i - m(\mathbf{X}_i)) \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{x})\}}}{n\mu_n(A_n(\mathbf{x}))} \right)^2 \middle| \mathbf{X}_1, \dots, \mathbf{X}_n \right\} \\ &= \frac{\sum_{i=1}^n \text{Var}(Y_i | \mathbf{X}_i) \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{x})\}}}{(n\mu_n(A_n(\mathbf{x})))^2} \\ &\leq \frac{\sigma^2}{n\mu_n(A_n(\mathbf{x}))} \mathbb{I}_{\{n\mu_n(A_n(\mathbf{x})) > 0\}}. \end{aligned}$$

Másrészt a torzítás típusú tagra a Jensen egyenlőtlenségből következik, hogy

$$\begin{aligned}
(\hat{m}_n(\mathbf{x}) - m(\mathbf{x}))^2 &= \left(\frac{\sum_{i=1}^n (m(\mathbf{X}_i) - m(\mathbf{x})) \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{x})\}}}{n\mu_n(A_n(\mathbf{x}))} \right)^2 \mathbb{I}_{\{n\mu_n(A_n(\mathbf{x})) > 0\}} \\
&\quad + m(\mathbf{x})^2 \mathbb{I}_{\{n\mu_n(A_n(\mathbf{x})) = 0\}} \\
&\leq \frac{\sum_{i=1}^n (m(\mathbf{X}_i) - m(\mathbf{x}))^2 \mathbb{I}_{\{\mathbf{X}_i \in A_n(\mathbf{x})\}}}{n\mu_n(A_n(\mathbf{x}))} \mathbb{I}_{\{n\mu_n(A_n(\mathbf{x})) > 0\}} \\
&\quad + m(\mathbf{x})^2 \mathbb{I}_{\{n\mu_n(A_n(\mathbf{x})) = 0\}} \\
&\leq d \cdot C^2 h_n^2 \mathbb{I}_{\{n\mu_n(A_n(\mathbf{x})) > 0\}} + m(\mathbf{x})^2 \mathbb{I}_{\{n\mu_n(A_n(\mathbf{x})) = 0\}} \\
&\quad \text{(a (2.3) és a } \max_{\mathbf{z} \in A_n(\mathbf{x})} \|\mathbf{x} - \mathbf{z}\|^2 \leq d \cdot h_n^2 \text{ miatt)} \\
&\leq d \cdot C^2 h_n^2 + m(\mathbf{x})^2 \mathbb{I}_{\{n\mu_n(A_n(\mathbf{x})) = 0\}}.
\end{aligned}$$

Az általánosság rovása nélkül feltehetjük, hogy a kompakt tartó S maga is egy kocka, mégpedig az $A_{n,1}, \dots, A_{n,l_n}$ cellák uniója. Akkor

$$l_n \leq \frac{\tilde{c}}{h_n^d}$$

valamilyen \tilde{c} konstansra, amelyik arányos az S térfogatával, tehát a 2.1 lemma és a (2.4) miatt

$$\begin{aligned}
&\mathbb{E} \left\{ \int (m_n(\mathbf{x}) - m(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} \\
&= \mathbb{E} \left\{ \int (m_n(\mathbf{x}) - \hat{m}_n(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} + \mathbb{E} \left\{ \int (\hat{m}_n(\mathbf{x}) - m(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} \\
&= \sum_{j=1}^{l_n} \mathbb{E} \left\{ \int_{A_{n,j}} (m_n(\mathbf{x}) - \hat{m}_n(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} \\
&\quad + \sum_{j=1}^{l_n} \mathbb{E} \left\{ \int_{A_{n,j}} (\hat{m}_n(\mathbf{x}) - m(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} \\
&\leq \sum_{j=1}^{l_n} \mathbb{E} \left\{ \frac{\sigma^2 \mu(A_{n,j})}{n\mu_n(A_{n,j})} \mathbb{I}_{\{\mu_n(A_{n,j}) > 0\}} \right\} + dC^2 h_n^2 \\
&\quad + \sum_{j=1}^{l_n} \mathbb{E} \left\{ \int_{A_{n,j}} m(\mathbf{x})^2 \mu(d\mathbf{x}) \mathbb{I}_{\{\mu_n(A_{n,j}) = 0\}} \right\},
\end{aligned}$$

tehát

$$\begin{aligned}
& \mathbb{E} \left\{ \int (m_n(\mathbf{x}) - m(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} \\
& \leq \sum_{j=1}^{l_n} \frac{2\sigma^2 \mu(A_{n,j})}{n\mu(A_{n,j})} + dC^2 h_n^2 + \sum_{j=1}^{l_n} \int_{A_{n,j}} m(\mathbf{x})^2 \mu(d\mathbf{x}) \mathbb{P}\{\mu_n(A_{n,j}) = 0\} \\
& \leq l_n \frac{2\sigma^2}{n} + dC^2 h_n^2 + \sup_{\mathbf{z} \in S} \{m(\mathbf{z})^2\} \sum_{j=1}^{l_n} \mu(A_{n,j}) (1 - \mu(A_{n,j}))^n \\
& \leq l_n \frac{2\sigma^2}{n} + dC^2 h_n^2 + l_n \frac{\sup_{\mathbf{z} \in S} m(\mathbf{z})^2}{n} \sup_j n\mu(A_{n,j}) e^{-n\mu(A_{n,j})} \\
& \leq l_n \frac{2\sigma^2}{n} + dC^2 h_n^2 + l_n \frac{\sup_{\mathbf{z} \in S} m(\mathbf{z})^2 e^{-1}}{n} \\
& \quad \text{(mivel } \sup_z z e^{-z} = e^{-1}) \\
& \leq \frac{(2\sigma^2 + \sup_{\mathbf{z} \in S} m(\mathbf{z})^2 e^{-1}) \tilde{c}}{n h_n^d} + dC^2 h_n^2.
\end{aligned}$$

□

3. fejezet

A magfüggvényes becslés

3.1. Bevezetés

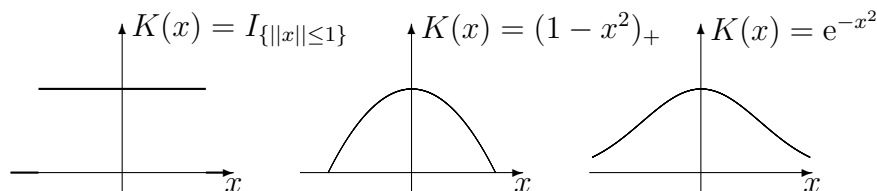
A magfüggvényes regresszióbecslés a következő alakú:

$$m_n(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)},$$

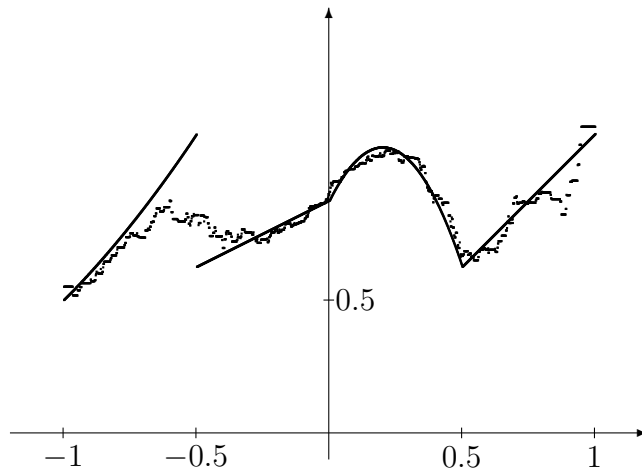
amennyiben a nevező pozitív, és 0 egyébként. Itt a $h_n > 0$ sávszélesség csak az n mintanagyságtól függ, míg a $K : \mathbb{R}^d \rightarrow [0, \infty)$ függvényt magfüggvénynek hívjuk. A 3.1 ábra mutat néhány példát. Általában $K(\mathbf{x})$ „nagy”, ha $\|\mathbf{x}\|$ „kicsi,” ezért a magfüggvényes becslés is lokális átlagolás típusú.

A 3.2–3.5 ábrák a magfüggvényes becslés szemléltetik naiv magfüggvény ($K(\mathbf{x}) = \mathbb{I}_{\{\|\mathbf{x}\| \leq 1\}}$) és Epanechnikov magfüggvény ($K(\mathbf{x}) = (1 - \|\mathbf{x}\|^2)_+$) és különböző h_n választás esetén az 1. fejezetben leírt szimulált adatokon.

A 3.6 ábra mutatja az L_2 hibát h függvényeként.



3.1. ábra. Példák magfüggvényre.

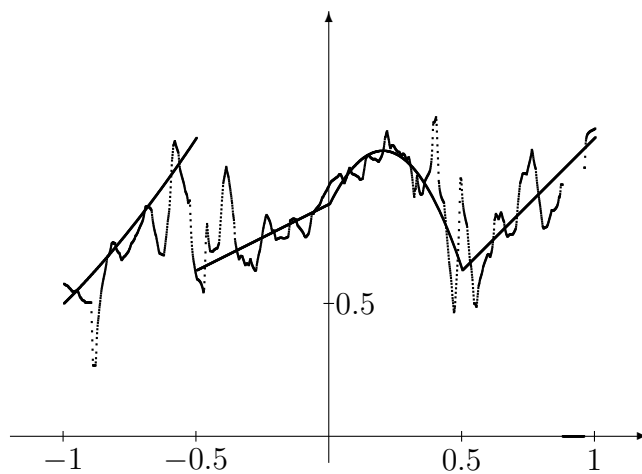


3.2. ábra. Magfüggvényes becslés naiv magfüggvénnyel: $h = 0.1$, L_2 hiba = 0.004066.

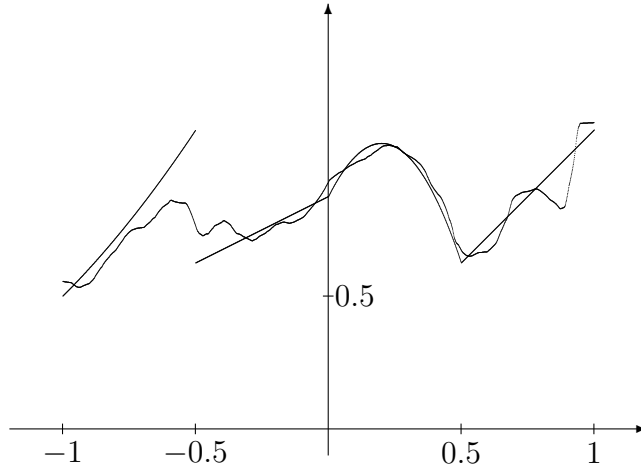
3.2. Konzisztencia

Ebben a szakaszban ismét a Stone tétel segítségével bizonyítjuk a konzisztenciát.

3.2.1. tétel *Tegyük fel, hogy létezik két origó közepű gömb $S_{0,r}$ r sugárral és $S_{0,R}$ R*



3.3. ábra. Magfüggvényes becslés Epanechnikov magfüggvénnyel: $h = 0.03$, L_2 hiba = 0.031560.



3.4. ábra. Magfüggvényes becslés Epanechnikov magfüggvénnyel: $h = 0.1$, L_2 hiba = 0.003608.

sugárral ($0 < r \leq R$) és egy $b > 0$ konstans úgy, hogy

$$\mathbb{I}_{\{\mathbf{x} \in S_{0,R}\}} \geq K(\mathbf{x}) \geq b \mathbb{I}_{\{\mathbf{x} \in S_{0,r}\}}$$

(dobozos magfüggvény). Legyen m_n a magfüggvényes becslés. Ha $h_n \rightarrow 0$ és $nh_n^d \rightarrow \infty$, akkor a magfüggvényes becslés gyengén univerzálisan konzisztens.

A 3.7 ábra is mutatja, hogy a dobozos magfüggvény kompakt tartójú, és az origó egy környezetében pozitív alsó korlátja van.

BIZONYÍTÁS. Legyen

$$K_h(\mathbf{x}) = K(\mathbf{x}/h).$$

Ellenőrizzük a 2.2.1. tétel feltételeit, amikor a súlyok

$$W_{n,i}(\mathbf{x}) = \frac{K_h(\mathbf{x} - \mathbf{X}_i)}{\sum_{j=1}^n K_h(\mathbf{x} - \mathbf{X}_j)}.$$

Az (i) feltétel azt jelenti, hogy

$$\mathbb{E} \left\{ \frac{\sum_{i=1}^n K_h(\mathbf{X} - \mathbf{X}_i) f(\mathbf{X}_i)}{\sum_{j=1}^n K_h(\mathbf{X} - \mathbf{X}_j)} \right\} \leq c \mathbb{E}\{f(\mathbf{X})\}$$

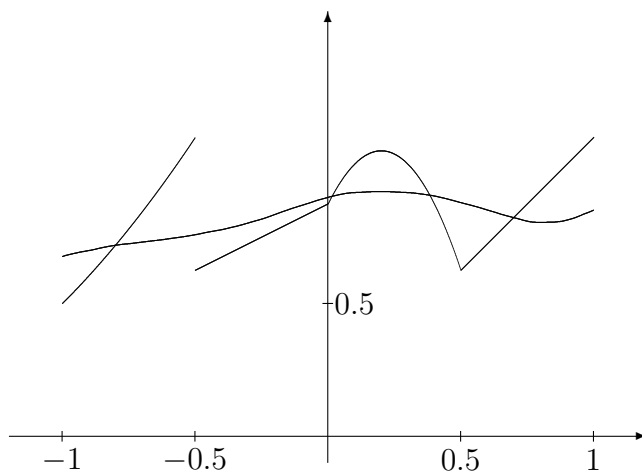
egy $c > 0$ konstanssal. A

$$\begin{aligned}
& \mathbb{E} \left\{ \frac{\sum_{i=1}^n K_h(\mathbf{X} - \mathbf{X}_i) f(\mathbf{X}_i)}{\sum_{j=1}^n K_h(\mathbf{X} - \mathbf{X}_j)} \right\} \\
&= n \mathbb{E} \left\{ \frac{K_h(\mathbf{X} - \mathbf{X}_1) f(\mathbf{X}_1)}{\sum_{j=1}^n K_h(\mathbf{X} - \mathbf{X}_j)} \right\} \\
&= n \mathbb{E} \left\{ \frac{K_h(\mathbf{X} - \mathbf{X}_1) f(\mathbf{X}_1)}{K_h(\mathbf{X} - \mathbf{X}_1) + \sum_{j=2}^n K_h(\mathbf{X} - \mathbf{X}_j)} \right\} \\
&= n \int f(\mathbf{u}) \left[\mathbb{E} \left\{ \int \frac{K_h(\mathbf{x} - \mathbf{u})}{K_h(\mathbf{x} - \mathbf{u}) + \sum_{j=2}^n K_h(\mathbf{x} - \mathbf{X}_j)} \mu(d\mathbf{x}) \right\} \right] \mu(d\mathbf{u})
\end{aligned}$$

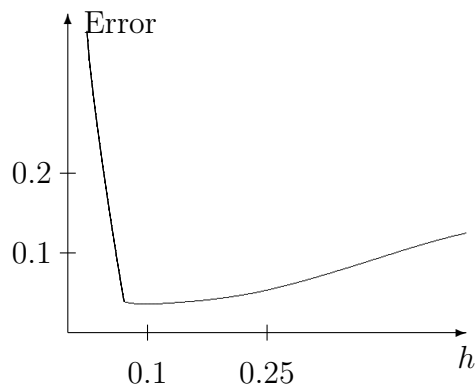
miatt elegendő megmutatni, hogy minden \mathbf{u} -ra és n -re teljesül, hogy

$$\mathbb{E} \left\{ \int \frac{K_h(\mathbf{x} - \mathbf{u})}{K_h(\mathbf{x} - \mathbf{u}) + \sum_{j=2}^n K_h(\mathbf{x} - \mathbf{X}_j)} \mu(d\mathbf{x}) \right\} \leq \frac{c}{n}.$$

A K kompakt tartóját lefedhetjük véges sok gömbbel, amelyek az $S_{0,r/2}$ gömb eltoltjai, ahol $r > 0$ az a sugar, amelyik a dobozos magfüggvény definíciójában szerepel. Jelölje



3.5. ábra. Magfüggvényes becslés Epanechnikov magfüggvénnyel: $h = 0.5$, L_2 hiba = 0.012551.



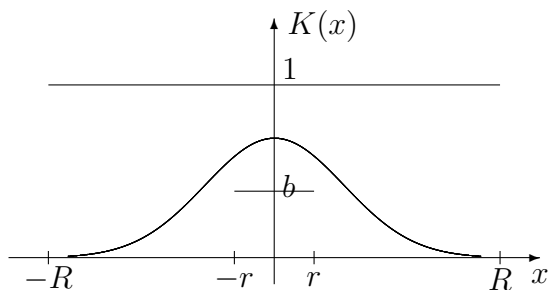
3.6. ábra. Az L_2 hiba Epanechnikov magfüggvény esetén mint a h függvénye.

$\mathbf{x}_i, i = 1, 2, \dots, M$ az eltolt gömbök közepeit! Ekkor minden \mathbf{x} -re és \mathbf{u} -ra

$$K_h(\mathbf{x} - \mathbf{u}) \leq \sum_{k=1}^M \mathbb{I}_{\{\mathbf{x} \in \mathbf{u} + h\mathbf{x}_k + S_{0, rh/2}\}}.$$

Továbbá $\mathbf{x} \in \mathbf{u} + h\mathbf{x}_k + S_{0, rh/2}$ -ből következik, hogy

$$\mathbf{u} + h\mathbf{x}_k + S_{0, rh/2} \subset \mathbf{x} + S_{0, rh}$$



3.7. ábra. Doboos magfüggvény.

r

$$\mathbf{x} \quad \mathbf{z} \quad \frac{r}{2}$$

3.8. ábra. Ha $\mathbf{x} \in S_{\mathbf{z}, r/2}$, akkor $S_{\mathbf{z}, r/2} \subseteq S_{\mathbf{x}, r}$.

(lásd a 3.8 ábrát). Ebből a két egyenlőtlenségből azt kapjuk, hogy

$$\begin{aligned} & \mathbb{E} \left\{ \int \frac{K_h(\mathbf{x} - \mathbf{u})}{K_h(\mathbf{x} - \mathbf{u}) + \sum_{j=2}^n K_h(\mathbf{x} - \mathbf{X}_j)} \mu(d\mathbf{x}) \right\} \\ & \leq \sum_{k=1}^M \mathbb{E} \left\{ \int_{\mathbf{u} + h\mathbf{x}_k + S_{0, rh/2}} \frac{K_h(\mathbf{x} - \mathbf{u})}{K_h(\mathbf{x} - \mathbf{u}) + \sum_{j=2}^n K_h(\mathbf{x} - \mathbf{X}_j)} \mu(d\mathbf{x}) \right\} \\ & \leq \sum_{k=1}^M \mathbb{E} \left\{ \int_{\mathbf{u} + h\mathbf{x}_k + S_{0, rh/2}} \frac{1}{1 + \sum_{j=2}^n K_h(\mathbf{x} - \mathbf{X}_j)} \mu(d\mathbf{x}) \right\} \\ & \leq \frac{1}{b} \sum_{k=1}^M \mathbb{E} \left\{ \int_{\mathbf{u} + h\mathbf{x}_k + S_{0, rh/2}} \frac{1}{1 + \sum_{j=2}^n \mathbb{I}_{\{\mathbf{X}_j \in \mathbf{x} + S_{0, rh}\}}} \mu(d\mathbf{x}) \right\} \\ & \leq \frac{1}{b} \sum_{k=1}^M \mathbb{E} \left\{ \int_{\mathbf{u} + h\mathbf{x}_k + S_{0, rh/2}} \frac{1}{1 + \sum_{j=2}^n \mathbb{I}_{\{\mathbf{X}_j \in \mathbf{u} + h\mathbf{x}_k + S_{0, rh/2}\}}} \mu(d\mathbf{x}) \right\} \\ & = \frac{1}{b} \sum_{k=1}^M \mathbb{E} \left\{ \frac{\mu(\mathbf{u} + h\mathbf{x}_k + S_{0, rh/2})}{1 + \sum_{j=2}^n \mathbb{I}_{\{\mathbf{X}_j \in \mathbf{u} + h\mathbf{x}_k + S_{0, rh/2}\}}} \right\} \\ & \leq \frac{1}{b} \sum_{k=1}^M \frac{\mu(\mathbf{u} + h\mathbf{x}_k + S_{0, rh/2})}{n\mu(\mathbf{u} + h\mathbf{x}_k + S_{0, rh/2})} \\ & \quad (\text{a 2.1 lemma miatt}) \\ & \leq \frac{M}{nb}. \end{aligned}$$

Az (ii) feltétel teljesül $D = 1$ konstanssal.

Az (iii) feltétel teljesül, mivel $h_n R < a$ esetén

$$\sum_{i=1}^n |W_{n,i}(\mathbf{X})| \mathbb{I}_{\{\|\mathbf{x}_i - \mathbf{x}\| > a\}} = \frac{\sum_{i=1}^n K_{h_n}(\mathbf{X} - \mathbf{X}_i) \mathbb{I}_{\{\|\mathbf{x}_i - \mathbf{x}\| > a\}}}{\sum_{i=1}^n K_{h_n}(\mathbf{X} - \mathbf{X}_i)} = 0.$$

Az (iv) feltétellel kapcsolatban megjegyezzük, hogy

$$1 - \sum_{i=1}^n W_{n,i}(\mathbf{X}) = \mathbb{I}_{\{\sum_{i=1}^n K_{h_n}(\mathbf{X} - \mathbf{X}_i) = 0\}},$$

ezért

$$\begin{aligned} \mathbb{P} \left\{ 1 \neq \sum_{i=1}^n W_{n,i}(\mathbf{X}) \right\} &= \mathbb{P} \left\{ \sum_{i=1}^n K_{h_n}(\mathbf{X} - \mathbf{X}_i) = 0 \right\} \\ &\leq \mathbb{P} \left\{ \sum_{i=1}^n \mathbb{I}_{\{\mathbf{x}_i \notin S_{\mathbf{x}, rh_n}\}} = 0 \right\} \\ &= \mathbb{P} \{ \mu_n(S_{\mathbf{x}, rh_n}) = 0 \} \\ &= \int (1 - \mu(S_{\mathbf{x}, rh_n}))^n \mu(d\mathbf{x}). \end{aligned}$$

Válasszunk egy origó közepű S gömböt, akkor

$$\begin{aligned} &\mathbb{P} \left\{ 1 \neq \sum_{i=1}^n W_{n,i}(\mathbf{X}) \right\} \\ &\leq \int_S e^{-n\mu(S_{\mathbf{x}, rh_n})} \mu(d\mathbf{x}) + \mu(S^c) \\ &= \int_S n\mu(S_{\mathbf{x}, rh_n}) e^{-n\mu(S_{\mathbf{x}, rh_n})} \frac{1}{n\mu(S_{\mathbf{x}, rh_n})} \mu(d\mathbf{x}) + \mu(S^c) \\ &= \max_u u e^{-u} \int_S \frac{1}{n\mu(S_{\mathbf{x}, rh_n})} \mu(d\mathbf{x}) + \mu(S^c). \end{aligned}$$

Az S gömb választásával a második tag tetszőlegesen kicsi lehet. Az első tag esetén található $\mathbf{z}_1, \dots, \mathbf{z}_{M_n}$ pontok úgy, hogy az $S_{\mathbf{z}_1, rh_n/2}, \dots, S_{\mathbf{z}_{M_n}, rh_n/2}$ gömbök uniója lefedi S -et, és

$$M_n \leq \frac{\tilde{c}}{h_n^d}.$$

Ekkor

$$\begin{aligned}
\int_S \frac{1}{n\mu(S_{\mathbf{x},rh_n})} \mu(d\mathbf{x}) &\leq \sum_{j=1}^{M_n} \int \frac{\mathbb{I}_{\{\mathbf{x} \in S_{\mathbf{z}_j, rh_n/2}\}}}{n\mu(S_{\mathbf{x},rh_n})} \mu(d\mathbf{x}) \\
&\leq \sum_{j=1}^{M_n} \int \frac{\mathbb{I}_{\{\mathbf{x} \in S_{\mathbf{z}_j, rh_n/2}\}}}{n\mu(S_{\mathbf{z}_j, rh_n/2})} \mu(d\mathbf{x}) \\
&\leq \frac{M_n}{n} \\
&\leq \frac{\tilde{c}}{nh_n^d} \rightarrow 0.
\end{aligned} \tag{3.1}$$

A (v) feltétel igazolása is egyszerű, mivel $K(\mathbf{x}) \leq 1$, és így minden $\delta > 0$ -ra

$$\begin{aligned}
\sum_{i=1}^n W_{n,i}(\mathbf{X})^2 &= \frac{\sum_{i=1}^n K_{h_n}(\mathbf{X} - \mathbf{X}_i)^2}{\left(\sum_{i=1}^n K_{h_n}(\mathbf{X} - \mathbf{X}_i)\right)^2} \\
&\leq \frac{\sum_{i=1}^n K_{h_n}(\mathbf{X} - \mathbf{X}_i)}{\left(\sum_{i=1}^n K_{h_n}(\mathbf{X} - \mathbf{X}_i)\right)^2} \\
&\leq \min \left\{ \delta, \frac{1}{\sum_{i=1}^n K_{h_n}(\mathbf{X} - \mathbf{X}_i)} \right\} \\
&\leq \min \left\{ \delta, \frac{1}{\sum_{i=1}^n b \mathbb{I}_{\{\mathbf{X}_i \in S_{\mathbf{X}, rh_n}\}}} \right\} \\
&\leq \delta + \frac{1}{\sum_{i=1}^n b \mathbb{I}_{\{\mathbf{X}_i \in S_{\mathbf{X}, rh_n}\}}} \mathbb{I}_{\{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in S_{\mathbf{X}, rh_n}\}} > 0\}},
\end{aligned}$$

ezért elegendő megmutatni, hogy

$$\mathbb{E} \left\{ \frac{1}{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in S_{\mathbf{X}, rh_n}\}}} \mathbb{I}_{\{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in S_{\mathbf{X}, rh_n}\}} > 0\}} \right\} \rightarrow 0.$$

Az előbbi S gömbre

$$\begin{aligned}
& \mathbb{E} \left\{ \frac{1}{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in S_{\mathbf{X}, rh_n}\}}} \mathbb{I}_{\{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in S_{\mathbf{X}, rh_n}\}} > 0\}} \right\} \\
& \leq \mathbb{E} \left\{ \frac{1}{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in S_{\mathbf{X}, rh_n}\}}} \mathbb{I}_{\{\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in S_{\mathbf{X}, rh_n}\}} > 0\}} \mathbb{I}_{\{\mathbf{X} \in S\}} \right\} + \mu(S^c) \\
& \leq 2\mathbb{E} \left\{ \frac{1}{(n+1)\mu(S_{\mathbf{X}, h_n})} \mathbb{I}_{\{\mathbf{X} \in S\}} \right\} + \mu(S^c) \\
& \quad (\text{a 2.1 lemma miatt}) \\
& \rightarrow \mu(S^c),
\end{aligned}$$

és így az S gömb választásával a bizonyítás teljes. \square

3.3. A konvergencia sebessége

Ebben a szakaszban kiszámoljuk a $\mathbb{E}\|m_n - m\|^2$ konvergencia-sebességét naiv magfüggvény és Lipschitz folytonos regressziófüggvény esetén.

3.3.2. tétel *Naiv magfüggvény és magfüggvényes becslés esetén tegyük fel, hogy*

$$\text{Var}(Y|\mathbf{X} = \mathbf{x}) \leq \sigma^2, \quad \mathbf{x} \in \mathbb{R}^d,$$

és

$$|m(\mathbf{x}) - m(\mathbf{z})| \leq C\|\mathbf{x} - \mathbf{z}\|, \quad \mathbf{x}, \mathbf{z} \in \mathbb{R}^d,$$

továbbá \mathbf{X} tartója a kompakt S^* halmaz. Akkor

$$\mathbb{E}\|m_n - m\|^2 \leq \hat{c} \frac{\sigma^2 + \sup_{\mathbf{z} \in S^*} |m(\mathbf{z})|^2}{n \cdot h_n^d} + C^2 h_n^2,$$

ahol \hat{c} csak az S^* átmérőjétől és d -től függ, ezért

$$h_n = c' \left(\frac{\sigma^2 + \sup_{\mathbf{z} \in S^*} |m(\mathbf{z})|^2}{C^2} \right)^{1/(d+2)} n^{-\frac{1}{d+2}}$$

választás esetén

$$\mathbb{E}\|m_n - m\|^2 \leq c'' \left(\sigma^2 + \sup_{\mathbf{z} \in S^*} |m(\mathbf{z})|^2 \right)^{2/(d+2)} C^{2d/(d+2)} n^{-2/(d+2)}.$$

BIZONYÍTÁS. A 2.4.3. tétel bizonyításához hasonlóan járunk el. Legyen

$$\hat{m}_n(\mathbf{x}) = \frac{\sum_{i=1}^n m(\mathbf{X}_i) \mathbb{I}_{\{\mathbf{x}_i \in S_{\mathbf{x}, h_n}\}}}{n\mu_n(S_{\mathbf{x}, h_n})},$$

akkor megkapjuk a (2.4) felbontást. Ha $B_n(\mathbf{x}) = \{n\mu_n(S_{\mathbf{x}, h_n}) > 0\}$, akkor

$$\begin{aligned} & \mathbb{E}\{(m_n(\mathbf{x}) - \hat{m}_n(\mathbf{x}))^2 | \mathbf{X}_1, \dots, \mathbf{X}_n\} \\ = & \mathbb{E} \left\{ \left(\frac{\sum_{i=1}^n (Y_i - m(\mathbf{X}_i)) \mathbb{I}_{\{\mathbf{x}_i \in S_{\mathbf{x}, h_n}\}}}{n\mu_n(S_{\mathbf{x}, h_n})} \right)^2 | \mathbf{X}_1, \dots, \mathbf{X}_n \right\} \\ = & \frac{\sum_{i=1}^n \text{Var}(Y_i | \mathbf{X}_i) \mathbb{I}_{\{\mathbf{x}_i \in S_{\mathbf{x}, h_n}\}}}{(n\mu_n(S_{\mathbf{x}, h_n}))^2} \\ \leq & \frac{\sigma^2}{n\mu_n(S_{\mathbf{x}, h_n})} \mathbb{I}_{B_n(\mathbf{x})}. \end{aligned}$$

A Jensen egyenlőtlenség és Lipschitz feltétel miatt

$$\begin{aligned} & (\hat{m}_n(\mathbf{x}) - m(\mathbf{x}))^2 \\ = & \left(\frac{\sum_{i=1}^n (m(\mathbf{X}_i) - m(\mathbf{x})) \mathbb{I}_{\{\mathbf{x}_i \in S_{\mathbf{x}, h_n}\}}}{n\mu_n(S_{\mathbf{x}, h_n})} \right)^2 \mathbb{I}_{B_n(\mathbf{x})} + m(\mathbf{x})^2 \mathbb{I}_{B_n(\mathbf{x})^c} \\ \leq & \frac{\sum_{i=1}^n (m(\mathbf{X}_i) - m(\mathbf{x}))^2 \mathbb{I}_{\{\mathbf{x}_i \in S_{\mathbf{x}, h_n}\}}}{n\mu_n(S_{\mathbf{x}, h_n})} \mathbb{I}_{B_n(\mathbf{x})} + m(\mathbf{x})^2 \mathbb{I}_{B_n(\mathbf{x})^c} \\ \leq & C^2 h_n^2 \mathbb{I}_{B_n(\mathbf{x})} + m(\mathbf{x})^2 \mathbb{I}_{B_n(\mathbf{x})^c} \\ \leq & C^2 h_n^2 + m(\mathbf{x})^2 \mathbb{I}_{B_n(\mathbf{x})^c}. \end{aligned}$$

Ezt és a 2.1 lemmát felhasználva kapjuk, hogy

$$\begin{aligned}
& \mathbb{E} \left\{ \int (m_n(\mathbf{x}) - m(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} \\
= & \mathbb{E} \left\{ \int (m_n(\mathbf{x}) - \hat{m}_n(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} + \mathbb{E} \left\{ \int (\hat{m}_n(\mathbf{x}) - m(\mathbf{x}))^2 \mu(d\mathbf{x}) \right\} \\
\leq & \int_{S^*} \mathbb{E} \left\{ \frac{\sigma^2}{n\mu(S_{\mathbf{x},h_n})} \mathbb{I}_{\{\mu_n(S_{\mathbf{x},h_n}) > 0\}} \right\} \mu(d\mathbf{x}) + C^2 h_n^2 \\
& + \int_{S^*} \mathbb{E} \left\{ m(\mathbf{x})^2 \mathbb{I}_{\{\mu_n(S_{\mathbf{x},h_n}) = 0\}} \right\} \mu(d\mathbf{x}) \\
\leq & \int_{S^*} \frac{2\sigma^2}{n\mu(S_{\mathbf{x},h_n})} \mu(d\mathbf{x}) + C^2 h_n^2 + \int_{S^*} m(\mathbf{x})^2 (1 - \mu(S_{\mathbf{x},h_n}))^n \mu(d\mathbf{x}) \\
\leq & \int_{S^*} \frac{2\sigma^2}{n\mu(S_{\mathbf{x},h_n})} \mu(d\mathbf{x}) + C^2 h_n^2 + \sup_{z \in S^*} m(z)^2 \int_{S^*} e^{-n\mu(S_{\mathbf{x},h_n})} \mu(d\mathbf{x}) \\
\leq & 2\sigma^2 \int_{S^*} \frac{1}{n\mu(S_{\mathbf{x},h_n})} \mu(d\mathbf{x}) + C^2 h_n^2 \\
& + \sup_{\mathbf{z} \in S^*} m(\mathbf{z})^2 \max_u u e^{-u} \int_{S^*} \frac{1}{n\mu(S_{\mathbf{x},h_n})} \mu(d\mathbf{x}).
\end{aligned}$$

A (3.1)-re hivatkozva úgy, hogy az ottani S gömb tartalmazza S^* -ot, ezekből az egyenlőségekből már következik a tétel állítása. \square

4. fejezet

k legközelebbi szomszéd becslés

4.1. Bevezetés

Rögzített $\mathbf{x} \in \mathbb{R}^d$ esetén rendezzük az $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ adatainkat az $\|\mathbf{X}_i - \mathbf{x}\|$ -ek növekvő sorrendjében. Az átrendezett adatainkat jelölje

$$(\mathbf{X}_{(1,n)}(\mathbf{x}), Y_{(1,n)}(\mathbf{x})), \dots, (\mathbf{X}_{(n,n)}(\mathbf{x}), Y_{(n,n)}(\mathbf{x}))$$

vagy egyszerűen

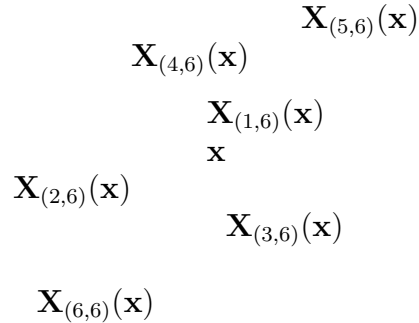
$$(\mathbf{X}_{(1,n)}, Y_{(1,n)}), \dots, (\mathbf{X}_{(n,n)}, Y_{(n,n)}),$$

ha nem származik belőle keveredés. $\mathbf{X}_{(k,n)}(\mathbf{x})$ -et az \mathbf{x} k -edik legközelebbi szomszédjának (k -NN) nevezzük.

A k_n -NN regresszióbecslést a következő módon definiáljuk:

$$m_n(\mathbf{x}) = \frac{1}{k_n} \sum_{i=1}^{k_n} Y_{(i,n)}(\mathbf{x}).$$

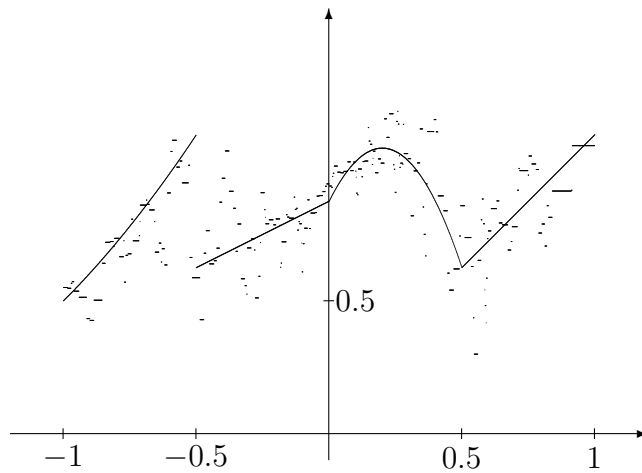
Ha \mathbf{X}_i és \mathbf{X}_j ugyanakkora távolságra vannak \mathbf{x} -től, azaz $\|\mathbf{X}_i - \mathbf{x}\| = \|\mathbf{X}_j - \mathbf{x}\|$, akkor fel kell oldani ezt a döntetlent. Ezt többféleképpen tehetjük meg. Például mondhatjuk, hogy \mathbf{X}_i „közelebb” van, ha $i < j$, azaz a döntetlent az indexek alapján oldjuk fel. Az egyszerűség kedvéért a későbbiekben feltesszük, hogy a döntetlen valószínűsége 0. Elvben ez a μ eloszlásra egy feltétel, például μ -nek nem lehet atomja, és így a szóbanforgó állítások nem univerzálisak, viszont az \mathbf{X} megfigyelési vektorhoz egy randomizáló komponenst hozzáadva automatikusan teljesül ez a feltétel. Formálisan, legyen (\mathbf{X}, Z) egy véletlen vektor úgy, hogy Z és (\mathbf{X}, Y) függetlenek, továbbá Z egyenletes eloszlású a $[0, 1]$



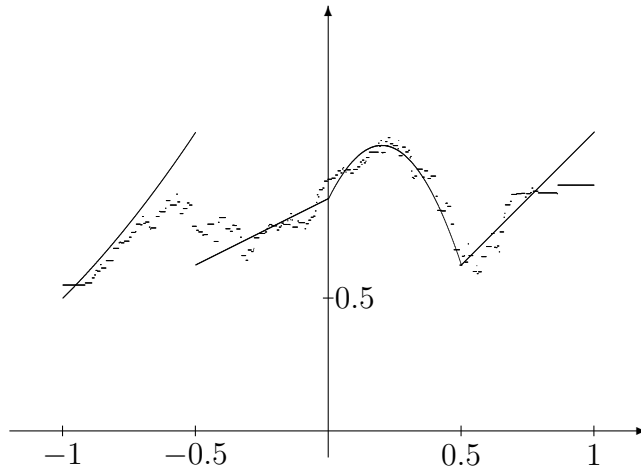
4.1. ábra. A legközelebbi szomszédok illusztrációja.

intervallumon. Ennek megfelelően kiegészítjük az adatainkat Z_1, Z_2, \dots, Z_n randomizálással úgy, hogy (\mathbf{X}_i, Z_i) és (\mathbf{X}, Z) azonos eloszlásúak. Ebben az esetben a döntetlen valószínűsége 0. A későbbiekben feltesszük, hogy az \mathbf{X} megfigyelésvektornak már van egy ilyen randomizáló komponense, és így módon minden \mathbf{x} -re az $\|\mathbf{X} - \mathbf{x}\|^2$ valószínűségi változó eloszlása abszolút folytonos.

A 4.2 – 4.4 ábrák a k_n -NN becslést illusztrálják a k_n különböző választásainál, illetve a 4.5 ábra mutatja az L_2 hibát a k_n függvényeként.



4.2. ábra. Kicsi $k_n = 3$, L_2 hiba = 0.011703.

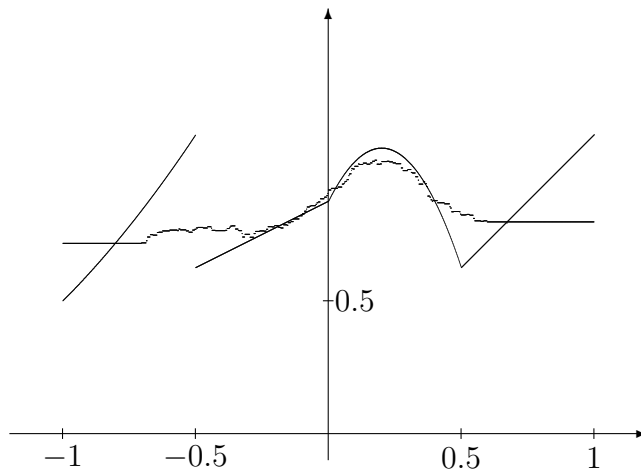


4.3. ábra. Jó $k_n = 12$, L_2 hiba = 0.004247.

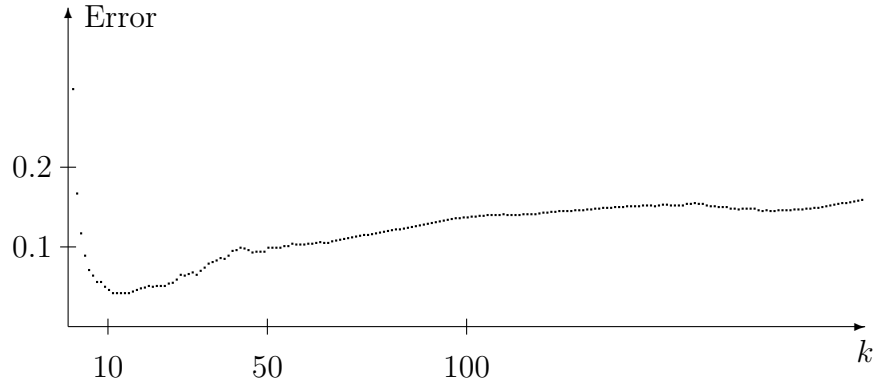
4.2. Konzisztencia

Ebben a szakaszban bizonyítjuk a k -NN becslés gyenge univerzális konzisztenciáját.

4.2.1. tétel *Ha $k_n \rightarrow \infty$, $k_n/n \rightarrow 0$, akkor a k_n -NN becslés gyengén konzisztens az (\mathbf{X}, Y) minden olyan eloszlására, ahol a döntetlen valószínűsége 0 és $\mathbb{E}Y^2 < \infty$.*



4.4. ábra. Nagy $k_n = 50$, L_2 error = 0.009931.



4.5. ábra. A k -NN becslés L_2 hibája a k függvényeként.

A Stone tétel feltételeinek az ellenőrzéséhez szükségünk van néhány lemmára.

4.1 lemma *Azon \mathbf{x} pontok halmazát, amelyekre $\mu(S_{\mathbf{x},\epsilon}) > 0$ minden $\epsilon > 0$ -ra, a μ tartójának hívjuk. Ha $\mathbf{x} \in \text{support}(\mu)$ és $\lim_{n \rightarrow \infty} k_n/n = 0$, akkor*

$$\|\mathbf{X}_{(k_n, n)}(\mathbf{x}) - \mathbf{x}\| \rightarrow 0$$

1 valószínűséggel.

BIZONYÍTÁS. Legyen $\epsilon > 0$. Definíció szerint $\mathbf{x} \in \text{support}(\mu)$ azt jelenti, hogy $\mu(S_{\mathbf{x},\epsilon}) > 0$. Vegyük észre, hogy

$$\{\|\mathbf{X}_{(k_n, n)}(\mathbf{x}) - \mathbf{x}\| > \epsilon\} = \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathbf{x}_i \in S_{\mathbf{x},\epsilon}\}} < \frac{k_n}{n} \right\}.$$

A nagy számok erős törvénye miatt

$$\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathbf{x}_i \in S_{\mathbf{x},\epsilon}\}} \rightarrow \mu(S_{\mathbf{x},\epsilon}) > 0$$

1 valószínűséggel, továbbá a lemma feltétele miatt

$$\frac{k_n}{n} \rightarrow 0.$$

Következésképp $\|\mathbf{X}_{(k_n, n)}(\mathbf{x}) - \mathbf{x}\| \rightarrow 0$ 1 valószínűséggel. □

4.2 lemma Legyen

$$B_a(\mathbf{x}') = \{\mathbf{x} : \mu(S_{\mathbf{x}, \|\mathbf{x}-\mathbf{x}'\|}) \leq a\}.$$

Akkor minden $\mathbf{x}' \in \mathcal{R}^d$ -ra

$$\mu(B_a(\mathbf{x}')) \leq \gamma_d a,$$

ahol γ_d csak d -től függ.

BIZONYÍTÁS. Legyen $C_j \subset \mathcal{R}^d$ egy $\pi/3$ szögű origó csúcsú kúp. Egy ilyen kúpnál ha $\mathbf{u}, \mathbf{u}' \in C_j$ és $\|\mathbf{u}\| < \|\mathbf{u}'\|$, akkor $\|\mathbf{u} - \mathbf{u}'\| < \|\mathbf{u}'\|$ (lásd 4.6 ábra). Legyen C_1, \dots, C_{γ_d} ilyen kúpoknak egy családja különböző irányokkal úgy, hogy az uniójuk lefedi \mathcal{R}^d -t:

$$\bigcup_{j=1}^{\gamma_d} C_j = \mathcal{R}^d.$$

Ekkor

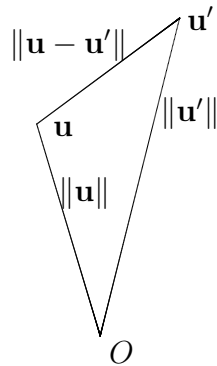
$$\mu(B_a(\mathbf{x}')) \leq \sum_{i=1}^{\gamma_d} \mu(\{\mathbf{x}' + C_i\} \cap B_a(\mathbf{x}')).$$

Legyen $\mathbf{x}^* \in \{\mathbf{x}' + C_i\} \cap B_a(\mathbf{x}')$. Akkor a kúp tulajdonság miatt azt kapjuk, hogy

$$\mu(\{\mathbf{x}' + C_i\} \cap S_{\mathbf{x}', \|\mathbf{x}' - \mathbf{x}^*\|} \cap B_a(\mathbf{x}')) \leq \mu(S_{\mathbf{x}^*, \|\mathbf{x}' - \mathbf{x}^*\|}) \leq a,$$

ahol kihasználtuk, hogy $\mathbf{x}^* \in B_a(\mathbf{x}')$. Mivel \mathbf{x}^* tetszőleges, ezért

$$\mu(\{\mathbf{x}' + C_i\} \cap B_a(\mathbf{x}')) \leq a,$$



4.6. ábra. A kúp tulajdonság.

és ezzel lemmát bebizonyítottuk. \square

A lemmának egy közvetkezménye, hogy azon $\mathbf{X}_1, \dots, \mathbf{X}_n$ pontok száma, amelyekre \mathbf{X} egyike az ő k legközelebbi szomszédja, nem nagyobb, mint egy konstansszor k :

4.1 következmény *Tegyük fel, hogy a döntetlen valószínűsége 0. Akkor*

$$\sum_{i=1}^n \mathbb{I}_{\{\mathbf{X} \text{ egyike az } \mathbf{X}_i \text{ k NN-jének } \{\mathbf{X}_1, \dots, \mathbf{X}_{i-1}, \mathbf{X}, \mathbf{X}_{i+1}, \dots, \mathbf{X}_n\}\text{-ből}\}} \leq k\gamma_d$$

1 valószínűséggel.

BIZONYÍTÁS. Alkalmazzuk a 4.2 lemmát $a = k/n$ és olyan μ esetén, amikor az az empirikus eloszlás μ_n a $\mathbf{X}_1, \dots, \mathbf{X}_n$ mintaára, azaz minden $A \subseteq \mathbb{R}^d$ halmazra $\mu_n(A) = (1/n) \sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in A\}}$. Akkor

$$B_{k/n}(\mathbf{X}) = \{\mathbf{x} : \mu_n(S_{\mathbf{x}, \|\mathbf{x} - \mathbf{x}\|}) \leq k/n\}$$

és

$$\begin{aligned} & \mathbf{X}_i \in B_{k/n}(\mathbf{X}) \\ \Leftrightarrow & \mu_n(S_{\mathbf{X}_i, \|\mathbf{X}_i - \mathbf{x}\|}) \leq k/n \\ \Leftrightarrow & \mathbf{X} \text{ egyike az } \mathbf{X}_i \text{ k NN-jének } \{\mathbf{X}_1, \dots, \mathbf{X}_{i-1}, \mathbf{X}, \mathbf{X}_{i+1}, \dots, \mathbf{X}_n\}\text{-ből} \end{aligned}$$

1 valószínűséggel, ahol a második \Leftrightarrow -nál alkalmaztuk azt a feltételt, hogy a döntetlen valószínűsége 0. Ez a 4.2 lemmával együtt azt eredményezi, hogy

$$\begin{aligned} & \sum_{i=1}^n \mathbb{I}_{\{\mathbf{X} \text{ egyike az } \mathbf{X}_i \text{ k NN-jének } \{\mathbf{X}_1, \dots, \mathbf{X}_{i-1}, \mathbf{X}, \mathbf{X}_{i+1}, \dots, \mathbf{X}_n\}\text{-ből}\}} \\ &= \sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in B_{k/n}(\mathbf{X})\}} \\ &= n \cdot \mu_n(B_{k/n}(\mathbf{X})) \\ &\leq k\gamma_d \end{aligned}$$

1 valószínűséggel. \square

4.3 lemma Tegyük fel, hogy a döntetlen valószínűsége 0. Akkor minden f integrálható függvényre és minden n -re és $k \leq n$ -ra

$$\sum_{i=1}^k \mathbb{E} \{ |f(\mathbf{X}_{(i,n)}(\mathbf{X}))| \} \leq k\gamma_d \mathbb{E} \{ |f(\mathbf{X})| \},$$

ahol γ_d csak a dimenziótól függ.

BIZONYÍTÁS. Ha f egy nemnegatív függvény, akkor a 4.1 következmény miatt

$$\begin{aligned} & \sum_{i=1}^k \mathbb{E} \{ f(\mathbf{X}_{(i,n)}(\mathbf{X})) \} \\ &= \mathbb{E} \left\{ \sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \text{ egyike az } \mathbf{X} \text{ k NN-jének } \{\mathbf{X}_1, \dots, \mathbf{X}_n\}\text{-ből}\}} f(\mathbf{X}_i) \right\} \\ &= \mathbb{E} \left\{ f(\mathbf{X}) \sum_{i=1}^n \mathbb{I}_{\{\mathbf{X} \text{ egyike az } \mathbf{X}_i \text{ k NN-jének } \{\mathbf{X}_1, \dots, \mathbf{X}_{i-1}, \mathbf{X}, \mathbf{X}_{i+1}, \dots, \mathbf{X}_n\}\text{-ből}\}} \right\} \\ & \quad (\text{felcserélve } \mathbf{X}\text{-et és } \mathbf{X}_i\text{-t}) \\ &\leq \mathbb{E} \{ f(\mathbf{X}) k\gamma_d \}, \end{aligned}$$

és ezzel a lemma bizonyítása kész. □

A 4.2.1. TÉTEL BIZONYÍTÁSA. Ismét a Stone tétel (2.2.1. tétel) feltételeit ellenőrizzük. A $W_{n,i}(\mathbf{X})$ súlyok értéke $1/k_n$, ha \mathbf{X}_i egyike az \mathbf{X} k_n legközelebbi szomszédjának, és 0 egyébként, tehát (ii) és (iv) automatikusan teljesül. A (v) feltétel is teljesül, mert $k_n \rightarrow \infty$. A (iii) feltételhez vegyük észre, hogy minden $\epsilon > 0$ -ra

$$\begin{aligned} & \mathbb{E} \left\{ \sum_{i=1}^n W_{n,i}(\mathbf{X}) \mathbb{I}_{\{\|\mathbf{X}_i - \mathbf{X}\| > \epsilon\}} \right\} \\ &= \int \mathbb{E} \left\{ \sum_{i=1}^n W_{n,i}(\mathbf{x}) \mathbb{I}_{\{\|\mathbf{X}_i - \mathbf{x}\| > \epsilon\}} \right\} \mu(d\mathbf{x}) \\ &= \int \mathbb{E} \left\{ \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{I}_{\{\|\mathbf{X}_{(i,n)}(\mathbf{x}) - \mathbf{x}\| > \epsilon\}} \right\} \mu(d\mathbf{x}) \rightarrow 0 \end{aligned}$$

teljesül, hacsak

$$\int \mathbb{P} \{ \|\mathbf{X}_{(k_n,n)}(\mathbf{x}) - \mathbf{x}\| > \epsilon \} \mu(d\mathbf{x}) \rightarrow 0, \quad (4.1)$$

ahol $\mathbf{X}_{(k_n, n)}(\mathbf{x})$ a \mathbf{x} jelöli k_n -edik legközelebbi szomszédját $\mathbf{X}_1, \dots, \mathbf{X}_n$ -ből. $\mathbf{x} \in \text{support}(\mu)$ esetén $k_n/n \rightarrow 0$ és 4.1 lemma miatt

$$\mathbb{P} \{ \|\mathbf{X}_{(k_n, n)}(\mathbf{x}) - \mathbf{x}\| > \epsilon \} \rightarrow 0 \quad (n \rightarrow \infty).$$

Ebből és a dominált konvergenciatételből következik (4.1). Végül az (i) feltételhez elég megmutatni, hogy minden nemnegatív f függvényre, amelyre $\mathbb{E}\{f(\mathbf{X})\} < \infty$, és minden n -re

$$\mathbb{E} \left\{ \sum_{i=1}^n \frac{1}{k_n} \mathbb{I}_{\{\mathbf{X}_i \text{ egyike az } \mathbf{x} \text{ k NN-jének } \{\mathbf{X}_1, \dots, \mathbf{X}_n\}\text{-ből}\}} f(\mathbf{X}_i) \right\} \leq c \cdot \mathbb{E} \{f(\mathbf{X})\}$$

valamilyen c konstansra. A 4.3 lemma miatt ez az egyenlőtlenség teljesül $c = \gamma_d$ konstanssal, és így a (i) feltételt is ellenőriztük. \square

4.3. A konvergencia-sebesség.

Ebben a szakaszban kiszámoljuk a $\mathbb{E}\|m_n - m\|^2$ konvergencia-sebességét a k_n -NN becslés esetén.

4.3.2. tétel *Tegyük fel, hogy \mathbf{X} korlátos,*

$$\sigma^2(\mathbf{x}) = \text{Var}(Y|\mathbf{X} = \mathbf{x}) \leq \sigma^2 \quad (\mathbf{x} \in \mathbb{R}^d)$$

és

$$|m(\mathbf{x}) - m(\mathbf{z})| \leq C\|\mathbf{x} - \mathbf{z}\| \quad (\mathbf{x}, \mathbf{z} \in \mathbb{R}^d),$$

továbbá $d \geq 3$. Legyen m_n a k_n -NN becslés. Akkor

$$\mathbb{E}\|m_n - m\|^2 \leq \frac{\sigma^2}{k_n} + c_1 \cdot C^2 \left(\frac{k_n}{n}\right)^{2/d},$$

ezért $k_n = c' (\sigma^2/C^2)^{d/(2+d)} n^{\frac{2}{d+2}}$ esetén

$$\mathbb{E}\|m_n - m\|^2 \leq c'' \sigma^{\frac{4}{d+2}} C^{\frac{2d}{2+d}} n^{-\frac{2}{d+2}}.$$

A 4.3.2. tétel bizonyításához szükségünk van a legközelebbi szomszédok távolságainak a konvergenciasebességére.

4.4 lemma *Tegyük fel, hogy \mathbf{X} korlátos. Ha $d \geq 3$, akkor*

$$\mathbb{E}\{\|\mathbf{X}_{(1,n)}(\mathbf{X}) - \mathbf{X}\|^2\} \leq \frac{\tilde{c}}{n^{2/d}}.$$

BIZONYÍTÁS. Rögzített $\epsilon > 0$ esetén

$$\mathbb{P}\{\|\mathbf{X}_{(1,n)}(\mathbf{X}) - \mathbf{X}\| > \epsilon\} = \mathbb{E}\{(1 - \mu(S_{\mathbf{x},\epsilon}))^n\}.$$

Legyen $A_1, \dots, A_{N(\epsilon)}$ a korlátos tartójú μ egy kockás partíciója úgy, hogy az A_j cellák átmérője ϵ , és

$$N(\epsilon) \leq \frac{c}{\epsilon^d}.$$

Ha $\mathbf{x} \in A_j$, akkor $A_j \subset S_{\mathbf{x},\epsilon}$, ezért

$$\begin{aligned} \mathbb{E}\{(1 - \mu(S_{\mathbf{x},\epsilon}))^n\} &= \sum_{j=1}^{N(\epsilon)} \int_{A_j} (1 - \mu(S_{\mathbf{x},\epsilon}))^n \mu(d\mathbf{x}) \\ &\leq \sum_{j=1}^{N(\epsilon)} \int_{A_j} (1 - \mu(A_j))^n \mu(d\mathbf{x}) \\ &= \sum_{j=1}^{N(\epsilon)} \mu(A_j) (1 - \mu(A_j))^n. \end{aligned}$$

Természetesen

$$\begin{aligned} \sum_{j=1}^{N(\epsilon)} \mu(A_j) (1 - \mu(A_j))^n &\leq \sum_{j=1}^{N(\epsilon)} \max_z z(1 - z)^n \\ &\leq \sum_{j=1}^{N(\epsilon)} \max_z z e^{-nz} \\ &= \frac{e^{-1} N(\epsilon)}{n}. \end{aligned}$$

Ha L jelöli a μ tartójának az átmérőjét, akkor

$$\begin{aligned}
\mathbb{E}\{\|\mathbf{X}_{(1,n)}(\mathbf{X}) - \mathbf{X}\|^2\} &= \int_0^\infty \mathbb{P}\{\|\mathbf{X}_{(1,n)}(\mathbf{X}) - \mathbf{X}\|^2 > \epsilon\} d\epsilon \\
&= \int_0^{L^2} \mathbb{P}\{\|\mathbf{X}_{(1,n)}(\mathbf{X}) - \mathbf{X}\| > \sqrt{\epsilon}\} d\epsilon \\
&\leq \int_0^{L^2} \min\left\{1, \frac{e^{-1}N(\sqrt{\epsilon})}{n}\right\} d\epsilon \\
&\leq \int_0^{L^2} \min\left\{1, \frac{c}{en}\epsilon^{-d/2}\right\} d\epsilon \\
&= \int_0^{(c/(en))^{2/d}} 1 d\epsilon + \frac{c}{en} \int_{(c/(en))^{2/d}}^{L^2} \epsilon^{-d/2} d\epsilon \\
&\leq \frac{\tilde{c}}{n^{2/d}}
\end{aligned}$$

$d \geq 3$ esetén. □

A 4.3.2. TÉTEL BIZONYÍTÁSA. Alkalmazzuk a következő felbontást:

$$\begin{aligned}
\mathbb{E}\{(m_n(\mathbf{x}) - m(\mathbf{x}))^2\} &= \mathbb{E}\{(m_n(\mathbf{x}) - \mathbb{E}\{m_n(\mathbf{x})|\mathbf{X}_1, \dots, \mathbf{X}_n\})^2\} \\
&\quad + \mathbb{E}\{(\mathbb{E}\{m_n(\mathbf{x})|\mathbf{X}_1, \dots, \mathbf{X}_n\} - m(\mathbf{x}))^2\} \\
&= I_1(\mathbf{x}) + I_2(\mathbf{x}).
\end{aligned}$$

Az első tag egyszerűbb:

$$\begin{aligned}
I_1(\mathbf{x}) &= \mathbb{E}\left\{\left(\frac{1}{k_n} \sum_{i=1}^{k_n} (Y_{(i,n)}(\mathbf{x}) - m(\mathbf{X}_{(i,n)}(\mathbf{x})))\right)^2\right\} \\
&= \mathbb{E}\left\{\frac{1}{k_n^2} \sum_{i=1}^{k_n} \sigma^2(\mathbf{X}_{(i,n)}(\mathbf{x}))\right\} \\
&\leq \frac{\sigma^2}{k_n}.
\end{aligned}$$

A második tag esetén

$$\begin{aligned}
I_2(\mathbf{x}) &= \mathbb{E} \left\{ \left(\frac{1}{k_n} \sum_{i=1}^{k_n} (m(\mathbf{X}_{(i,n)}(\mathbf{x})) - m(\mathbf{x})) \right)^2 \right\} \\
&\leq \mathbb{E} \left\{ \left(\frac{1}{k_n} \sum_{i=1}^{k_n} |m(\mathbf{X}_{(i,n)}(\mathbf{x})) - m(\mathbf{x})| \right)^2 \right\} \\
&\leq \mathbb{E} \left\{ \left(\frac{1}{k_n} \sum_{i=1}^{k_n} C \|\mathbf{X}_{(i,n)}(\mathbf{x}) - \mathbf{x}\| \right)^2 \right\}.
\end{aligned}$$

Legyen $N = k_n \lfloor \frac{n}{k_n} \rfloor$. Szegmentáljuk a $\mathbf{X}_1, \dots, \mathbf{X}_n$ adatainkat $k_n + 1$ részre úgy, hogy az első k_n rész hossza $\lfloor \frac{n}{k_n} \rfloor$, és legyen $\tilde{\mathbf{X}}_j^{\mathbf{x}}$ az \mathbf{x} első szomszédja a j -edik részből. Akkor $\tilde{\mathbf{X}}_1^{\mathbf{x}}, \dots, \tilde{\mathbf{X}}_{k_n}^{\mathbf{x}}$ a $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ -nek k_n darab különböző eleme úgy, hogy

$$\sum_{i=1}^{k_n} \|\mathbf{X}_{(i,n)}(\mathbf{x}) - \mathbf{x}\| \leq \sum_{j=1}^{k_n} \|\tilde{\mathbf{X}}_j^{\mathbf{x}} - \mathbf{x}\|,$$

tehát a Jensen egyenlőtlenség miatt

$$\begin{aligned}
I_2(\mathbf{x}) &\leq C^2 \mathbb{E} \left\{ \left(\frac{1}{k_n} \sum_{j=1}^{k_n} \|\tilde{\mathbf{X}}_j^{\mathbf{x}} - \mathbf{x}\| \right)^2 \right\} \\
&\leq C^2 \frac{1}{k_n} \sum_{j=1}^{k_n} \mathbb{E} \left\{ \|\tilde{\mathbf{X}}_j^{\mathbf{x}} - \mathbf{x}\|^2 \right\} \\
&= C^2 \mathbb{E} \left\{ \|\tilde{\mathbf{X}}_1^{\mathbf{x}} - \mathbf{x}\|^2 \right\} \\
&= C^2 \mathbb{E} \left\{ \|\mathbf{X}_{(1, \lfloor \frac{n}{k_n} \rfloor)}(\mathbf{x}) - \mathbf{x}\|^2 \right\}.
\end{aligned}$$

A 4.4 lemmából következik, hogy

$$\begin{aligned}
\frac{1}{C^2} \left\lfloor \frac{n}{k_n} \right\rfloor^{2/d} \int I_2(\mathbf{x}) \mu(d\mathbf{x}) &\leq \left\lfloor \frac{n}{k_n} \right\rfloor^{2/d} \mathbb{E} \left\{ \|\mathbf{X}_{(1, \lfloor \frac{n}{k_n} \rfloor)}(\mathbf{X}) - \mathbf{X}\|^2 \right\} \\
&\leq \text{const.}
\end{aligned}$$

□

5. fejezet

Idősorok predikciója

5.1. A predikciós probléma négyzetes hibával

Ebben a fejezetben valós értékű sorozatok predikcióját vizsgáljuk. Minden $t = 1, 2, \dots$ időpontban a prediktor ad egy becslést a valós értékű y_1, y_2, \dots sorozat t -edik elemére, azaz y_t -re, ha adott a sorozat $y_1^{t-1} = (y_1, \dots, y_{t-1})$ múltja (itt y_1^0 jelöli az üres szegmenst) és $\mathbf{x}_1^t = (\mathbf{x}_1, \dots, \mathbf{x}_t)$ megfigyelési vektorok, ahol $\mathbf{x}_t \in \mathbb{R}^d$. A prediktor becslése tehát a t időpontban az \mathbf{x}_1^t és a y_1^{t-1} szegmenseket használhatja fel. Maga a predikciós stratégia

$$g_t : (\mathbb{R}^d)^t \times \mathbb{R}^{t-1} \rightarrow \mathbb{R}$$

függvényeknek egy $g = \{g_t\}_{t=1}^\infty$ sorozata úgy, hogy a predikció a t időpontban $g_t(\mathbf{x}_1^t, y_1^{t-1})$.

A későbbiekben feltesszük, hogy $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots$ egy $(\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), \dots$ valószínűségi változósorozat realizációi úgy, hogy a $\{(\mathbf{X}_n, Y_n)\}_{-\infty}^\infty$ sztochasztikus folyamat stacionárius és ergodikus.

Az n -edik időpontban a prediktort az *átlagos négyzetes hibával* minősítjük:

$$L_n(g) = \frac{1}{n} \sum_{t=1}^n (g_t(\mathbf{X}_1^t, Y_1^{t-1}) - Y_t)^2.$$

Az a cél, hogy nagy n -re legyen $L_n(g)$ kicsi.

Ebben a vonatkozásban ismert az elvi optimum, ugyanis Algoet [1994] megmutatta, hogy minden g predikciós stratégiára és stacionárius, ergodikus $\{(\mathbf{X}_n, Y_n)\}_{-\infty}^\infty$ folyamatra

$$\liminf_{n \rightarrow \infty} L_n(g) \geq L^* \quad 1 \text{ valószínűséggel,} \quad (5.1)$$

ahol

$$L^* = \mathbf{E}(Y_0 - \mathbf{E}Y_0 | \mathbf{X}_{-\infty}^0, Y_{-\infty}^{-1})^2$$

a legkisebb átlagos négyzetes hiba, amennyiben Y_0 -ra adunk becslést a teljes $\mathbf{X}_{-\infty}^0, Y_{-\infty}^{-1}$ végtelen múlt ismeretében.

Ennek az elvi határnak az eléréséhez viszont szükség van a $\{(\mathbf{X}_n, Y_n)\}_{-\infty}^{\infty}$ folyamat együttes eloszlásainak az ismeretére, tehát egy igen fontos probléma, hogy csak a folyamat múltját megfigyelve vajon elérhető-e ez az elvi optimum.

5.1 definíció *A g predikciós stratégiát a $\{(\mathbf{X}_n, Y_n)\}_{-\infty}^{\infty}$ stacionárius, ergodikus folyamatok egy \mathcal{C} osztályára nézve univerzálisan konzeisztensnek hívunk, ha az osztály minden folyamatára*

$$\lim_{n \rightarrow \infty} L_n(g) = L^* \quad 1 \text{ valószínűséggel.}$$

5.2. Univerzálisan konzisztens partíciós stratégia

A következő szakaszokban korlátos ergodikus idősorok esetén vezetünk be univerzálisan konzisztens prediktorokat. Feltesszük, hogy $|Y_0| < B$, és ismert egy B korlát.

A predikciós stratégiát úgy definiáljuk, hogy úgynevezett *elemi prediktork* konvex lineáris kombinációját vesszük, ahol a súlyozó együtthatók függenek az elemi prediktork múltbeli hibáitól.

Bevezetünk elemi prediktork $h^{(k,\ell)}$, $k, \ell = 1, 2, \dots$ kétindexes halmazát a következő módon. Legyen $\mathcal{P}_\ell = \{A_{\ell,j}, j = 1, 2, \dots, m_\ell\}$ az \mathbb{R} partícióinak egy sorozata, és $\mathcal{Q}_\ell = \{B_{\ell,j}, j = 1, 2, \dots, m'_\ell\}$ az \mathbb{R}^d partícióinak egy sorozata. Ezek a partíciók generálnak kvantálódokat:

$$F_\ell(y) = j, \text{ ha } y \in A_{\ell,j}$$

és

$$G_\ell(\mathbf{x}) = j, \text{ ha } \mathbf{x} \in B_{\ell,j}.$$

Gyorsírással minden n -re és minden $y_1^n \in \mathbb{R}^n$ -re legyen $F_\ell(y_1^n)$ a $F_\ell(y_1), \dots, F_\ell(y_n)$ szegmens, és hasonlóan minden $\mathbf{x}_1^n \in (\mathbb{R}^d)^n$ -re legyen $G_\ell(\mathbf{x}_1^n)$ a $G_\ell(\mathbf{x}_1), \dots, G_\ell(\mathbf{x}_n)$ szegmens.

Rögzített k, ℓ természetes számok és minden $k+1$ hosszúságú, természetes számokból álló z szegmens és minden k hosszúságú, természetes számokból álló s szegmens esetén definiáljuk a partíciós regresszióbecslést:

$$\widehat{E}_n^{(k,\ell)}(\mathbf{x}_1^n, y_1^{n-1}, z, s) = \frac{\sum_{\{k < t < n : G_\ell(\mathbf{x}_{t-k}^t) = z, F_\ell(y_{t-k}^{t-1}) = s\}} y_t}{|\{k < t < n : G_\ell(\mathbf{x}_{t-k}^t) = z, F_\ell(y_{t-k}^{t-1}) = s\}|},$$

ha $n > k + 1$, ahol definíció szerint $0/0 = 0$.

Ekkor a $h^{(k,\ell)}$ elemi prediktor definíciója:

$$h_n^{(k,\ell)}(\mathbf{x}_1^n, y_1^{n-1}) = \widehat{E}_n^{(k,\ell)}(\mathbf{x}_1^n, y_1^{n-1}, G_\ell(\mathbf{x}_{n-k}^n), F_\ell(y_{n-k}^{n-1})),$$

azaz $h_n^{(k,\ell)}$ a \mathcal{Q}_ℓ és \mathcal{P}_ℓ partíciókkal kvantálja az $\mathbf{x}_1^n, y_1^{n-1}$ szegmenst, és megkeresi a közelmúlt $G_\ell(\mathbf{x}_{n-k}^n)$ és $F_\ell(y_{n-k}^{n-1})$ kvantált szegmenseihez hasonlókat a múltban, és a predikció ezen hasonló t időpontokhoz tartozó y_t -k átlaga.

A nemparaméteres regresszióbecslések elméletével ellentétben, ahol az adatok független, azonos eloszlásúak, ergodikus megfigyelések esetén lehetetlen megválasztani $k = k_n$ -et és $\ell = \ell_n$ -et úgy, hogy a kapott prediktor univerzálisan konzisztens legyen a korlátos, ergodikus folyamatok osztályára. A gépi tanulás újabb eredményei alapján viszont lehetséges ilyen elemi prediktorok aggregációja (részleteket lásd a Cesa-Bianchi and Lugosi [2006] könyvben). A javasolt predikciós stratégia a következő módon működik: válasszunk egy tetszőleges $\{q_{k,\ell}\}$ valószínűségeloszlást a (k, ℓ) párok felett úgy, hogy $q_{k,\ell} > 0$ minden k, ℓ -re. Legyen $c = 8B^2$, és vezessük be a

$$w_{t,k,\ell} = q_{k,\ell} e^{-(t-1)L_{t-1}(h^{(k,\ell)})/c} \quad (5.2)$$

súlyokat és azok normalizáltjait:

$$p_{t,k,\ell} = \frac{w_{t,k,\ell}}{W_t}, \quad (5.3)$$

ahol

$$W_t = \sum_{i,j=1}^{\infty} w_{t,i,j}. \quad (5.4)$$

A g kombinált predikciós stratégia az elemi prediktorok konvex lineáris kombinációja:

$$g_t(\mathbf{x}_1^t, y_1^{t-1}) = \sum_{k,\ell=1}^{\infty} p_{t,k,\ell} h^{(k,\ell)}(\mathbf{x}_1^t, y_1^{t-1}), \quad t = 1, 2, \dots \quad (5.5)$$

azaz egy elemi prediktornak akkor van nem elhanyagolható súlya az aggregálásban, ha a $t - 1$ időpontig jó a teljesítménye.

5.2.1. tétel (GYÖRFI AND LUGOSI [2002]) *Tegyük fel, hogy*

(a) *a \mathcal{P}_ℓ partíciók sorozata finomodó, azaz $\mathcal{P}_{\ell+1}$ minden cellája a \mathcal{P}_ℓ egy cellájának a részhalmaza, $\ell = 1, 2, \dots$;*

(b) *a \mathcal{Q}_ℓ partíciók sorozata finomodó;*

(c) a \mathcal{P}_ℓ partíciók sorozata aszimptotikusan finom, azaz minden origó közepű S gömbre

$$\lim_{\ell \rightarrow \infty} \max_{A \in \mathcal{P}_\ell, A \cap S \neq \emptyset} \text{diam}(A) = 0;$$

(d) a \mathcal{Q}_ℓ partíciók sorozata aszimptotikusan finom;

Akkor az előbb definiált g partíciós predikciós stratégia univerzálisan konzisztens a stacionárius, ergodikus $\{(\mathbf{X}_n, Y_n)\}_{-\infty}^{\infty}$ folyamatok azon osztályára, amelyre $|Y_0| \leq B$.

A tétel bizonyításának legfontosabb komponense a következő lemma, amelyet lényegében Kivinen és Warmuth [1999] bizonyított individuális sorozatokra.

5.1 lemma Legyen $\tilde{h}_1, \tilde{h}_2, \dots$ predikcióknak egy sorozata, és legyen $\{q_k\}$ egy eloszlás úgy, hogy $q_k > 0$ minden k természetes számra. Tegyük fel, hogy $\tilde{h}_i(\mathbf{x}_1^n, y_1^{n-1}) \in [-B, B]$ és $y_1^n \in [-B, B]^n$. Definiáljuk a

$$w_{t,k} = q_k e^{-(t-1)L_{t-1}(\tilde{h}_k)/c}$$

súlyokat, ahol $c \geq 8B^2$, és azok

$$v_{t,k} = \frac{w_{t,k}}{\sum_{i=1}^{\infty} w_{t,i}}$$

normalizáltját. Ha a \tilde{g} predikciós stratégiát a

$$\tilde{g}_t(\mathbf{x}_1^t, y_1^{t-1}) = \sum_{k=1}^{\infty} v_{t,k} \tilde{h}_k(\mathbf{x}_1^t, y_1^{t-1}) \quad t = 1, 2, \dots$$

egyenlőséggel definiáljuk, akkor minde $n \geq 1$ -re

$$L_n(\tilde{g}) \leq \inf_k \left(L_n(\tilde{h}_k) - \frac{c \ln q_k}{n} \right).$$

BIZONYÍTÁS. Legyen

$$W_1 = 1$$

és

$$W_t = \sum_{k=1}^{\infty} w_{t,k}$$

$t > 1$ esetén. Ekkor

$$W_{t+1} = \sum_{k=1}^{\infty} w_{t,k} e^{-(y_t - \tilde{h}_k(\mathbf{x}_1^t, y_1^{t-1}))^2/c} = W_t \sum_{k=1}^{\infty} v_{t,k} e^{-(y_t - \tilde{h}_k(\mathbf{x}_1^t, y_1^{t-1}))^2/c},$$

tehát

$$-c \ln \frac{W_{t+1}}{W_t} = -c \ln \left(\sum_{k=1}^{\infty} v_{t,k} e^{-(y_t - \tilde{h}_k(\mathbf{x}_1^t, y_1^{t-1}))^2 / c} \right).$$

Vezessük be az

$$F_t(z) = e^{-(y_t - z)^2 / c}$$

függvényt. Mivel $c \geq 8B^2$, ezért az F_t függvény konkáv a $[-B, B]$ intervallumon, tehát a Jensen egyenlőtlenség miatt

$$\left[\sum_{k=1}^{\infty} v_{t,k} \left(y_t - \tilde{h}_k(\mathbf{x}_1^t, y_1^{t-1}) \right) \right]^2 \leq -c \ln \frac{W_{t+1}}{W_t}, \quad (5.6)$$

következésképp

$$\begin{aligned} nL_n(\tilde{g}) &= \sum_{t=1}^n \left(y_t - \tilde{g}(\mathbf{x}_1^t, y_1^{t-1}) \right)^2 \\ &= \sum_{t=1}^n \left[\sum_{k=1}^{\infty} v_{t,k} \left(y_t - \tilde{h}_k(\mathbf{x}_1^t, y_1^{t-1}) \right) \right]^2 \\ &\leq -c \sum_{t=1}^n \ln \frac{W_{t+1}}{W_t} \\ &= -c \ln W_{n+1}, \end{aligned}$$

és így

$$\begin{aligned} nL_n(\tilde{g}) &\leq -c \ln \left(\sum_{k=1}^{\infty} w_{n+1,k} \right) \\ &= -c \ln \left(\sum_{k=1}^{\infty} q_k e^{-nL_n(\tilde{h}_k)/c} \right) \\ &\leq -c \ln \left(\sup_k q_k e^{-nL_n(\tilde{h}_k)/c} \right) \\ &= \inf_k \left(-c \ln q_k + nL_n(\tilde{h}_k) \right). \end{aligned}$$

□

A 5.2.1. tétel bizonyításának másik fontos komponense a Breiman féle általánosított ergodtétel, amelynek a bizonyítása megtalálható az Algoet [1994] cikkben vagy a Györfi et al. [2002] könyvben.

5.2 lemma (BREIMAN [1957]). *Legyen $Z = \{Z_i\}_{-\infty}^{\infty}$ egy stacionárius, ergodik folyamat. Jelölje T az eltolásoperátort. Legyen f_i valós értékű függvények egy sorozata úgy, hogy egy f függvényre $f_i(Z) \rightarrow f(Z)$ 1 valószínűséggel. Ha $\mathbb{E}\{\sup_i |f_i(Z)|\} < \infty$, akkor*

$$\lim_{t \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f_i(T^i Z) = \mathbb{E}\{f(Z)\} \quad 1 \text{ valószínűséggel.}$$

A 5.2.1. TÉTEL BIZONYÍTÁSA. (5.1) miatt elég azt megmutatni, hogy

$$\limsup_{n \rightarrow \infty} L_n(g) \leq L^*$$

1 valószínűséggel. Az ergodtétel kétszeri alkalmazásával kapjuk, hogy

$$\begin{aligned} \widehat{E}_n^{(k,\ell)}(\mathbf{X}_1^n, Y_1^{n-1}, z, s) &= \frac{\frac{1}{n} \sum_{\{k < i < n : G_\ell(\mathbf{X}_{t-k}^t) = z, F_\ell(Y_{t-k}^{t-1}) = s\}} Y_i}{\frac{1}{n} |\{k < i < n : G_\ell(\mathbf{X}_{t-k}^t) = z, F_\ell(Y_{t-k}^{t-1}) = s\}|} \\ &\rightarrow \frac{\mathbb{E}\{Y_0 \mathbb{I}_{\{G_\ell(\mathbf{X}_{-k}^0) = z, F_\ell(Y_{-k}^{-1}) = s\}}\}}{\mathbb{P}\{G_\ell(\mathbf{X}_{-k}^0) = z, F_\ell(Y_{-k}^{-1}) = s\}} \\ &= \mathbb{E}\{Y_0 | G_\ell(\mathbf{X}_{-k}^0) = z, F_\ell(Y_{-k}^{-1}) = s\}, \end{aligned}$$

1 valószínűséggel, ezért

$$\lim_{n \rightarrow \infty} \sup_z \sup_s |\widehat{E}_n^{(k,\ell)}(\mathbf{X}_1^n, Y_1^{n-1}, z, s) - \mathbb{E}\{Y_0 | G_\ell(\mathbf{X}_{-k}^0) = z, F_\ell(Y_{-k}^{-1}) = s\}| = 0$$

1 valószínűséggel. A 5.2 lemma miatt azt kapjuk, hogy

$$\begin{aligned} L_n(h^{(k,\ell)}) &= \frac{1}{n} \sum_{i=1}^n (h^{(k,\ell)}(\mathbf{X}_1^i, Y_1^{i-1}) - Y_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\widehat{E}_n^{(k,\ell)}(\mathbf{X}_1^i, Y_1^{i-1}, G_\ell(\mathbf{X}_{i-k}^i), F_\ell(Y_{i-k}^{i-1})) - Y_i)^2 \\ &\rightarrow \mathbb{E}\{(Y_0 - \mathbb{E}\{Y_0 | G_\ell(\mathbf{X}_{-k}^0), F_\ell(Y_{-k}^{-1})\})^2\} \\ &\stackrel{\text{def}}{=} \epsilon_{k,\ell}. \end{aligned}$$

1 valószínűséggel. A \mathcal{P}_ℓ és a \mathcal{Q}_ℓ partíciók finomodók, ezért $\mathbb{E}\{Y_0|G_\ell(\mathbf{X}_{-k}^0), F_\ell(Y_{-k}^{-1})\}$ a (k, ℓ) párral indexelt martingál. A martingál-konvergenciatételre (lásd Stout [1974]) és a (c), (d) feltételre hivatkozva azt kapjuk, hogy

$$\inf \epsilon_{k,\ell} = \lim_{k,\ell \rightarrow \infty} \epsilon_{k,\ell} = \mathbb{E}\left\{\left(Y_0 - \mathbb{E}\{Y_0|\mathbf{X}_{-\infty}^0, Y_{-\infty}^{-1}\}\right)^2\right\} = L^*.$$

Az 5.1 lemma miatt

$$L_n(g) \leq \inf_{k,\ell} \left(L_n(h^{(k,\ell)}) - \frac{c \ln q_{k,\ell}}{n} \right), \quad (5.7)$$

tehát

$$\begin{aligned} \limsup_{n \rightarrow \infty} L_n(g) &\leq \limsup_{n \rightarrow \infty} \inf_{k,\ell} \left(L_n(h^{(k,\ell)}) - \frac{c \ln q_{k,\ell}}{n} \right) \\ &\leq \inf_{k,\ell} \limsup_{n \rightarrow \infty} \left(L_n(h^{(k,\ell)}) - \frac{c \ln q_{k,\ell}}{n} \right) \\ &\leq \inf_{k,\ell} \limsup_{n \rightarrow \infty} L_n(h^{(k,\ell)}) \\ &= \inf_{k,\ell} \epsilon_{k,\ell} \\ &= \lim_{k,\ell \rightarrow \infty} \epsilon_{k,\ell} \\ &= L^* \end{aligned}$$

1 valószínűséggel. □

5.3. Univerzálisan konzisztens magfüggvényes stratégia

Ebben a szakaszban bevezetjük a *magfüggvényes* predikciót. Az egyszerűség kedvéért csak a naiv magfüggvény esetét tárgyaljuk. Megint elemi prediktorok egy kétindexes $h^{(k,\ell)}$ halmazát vezetjük be. Minden (k, ℓ) számpárhoz tartozik két sugár $r_{k,\ell} > 0$ és $r'_{k,\ell} > 0$ úgy, hogy minden rögzített k -ra

$$\lim_{\ell \rightarrow \infty} r_{k,\ell} = 0, \quad (5.8)$$

és

$$\lim_{\ell \rightarrow \infty} r'_{k,\ell} = 0. \quad (5.9)$$

Vezessük be az illeszkedések időpontjainak a halmazát:

$$J_n^{(k,\ell)} = \left\{ k < t < n : \|\mathbf{x}_{t-k}^t - \mathbf{x}_{n-k}^n\| \leq r_{k,\ell}, \|y_{t-k}^{t-1} - y_{n-k}^{n-1}\| \leq r'_{k,\ell} \right\}$$

Akkor az n időpontban a $h_n^{(k,\ell)}$ elemi prediktort a következő lokális átlagolással definiáljuk:

$$h_n^{(k,\ell)}(\mathbf{x}_1^n, y_1^{n-1}) = \frac{\sum_{\{t \in J_n^{(k,\ell)}\}} y_t}{|J_n^{(k,\ell)}|}, \quad n > k + 1, \quad (5.10)$$

ahol $0/0$ -t 0 -val definiáljuk. Ezeket az elemi prediktorokat ugyanúgy kombináljuk, mint a partíciós prediktornál (lásd (5.2), (5.3), (5.4) és (5.5)).

5.3.2. tétel *Tegyük fel (5.8)-t és (5.9)-t. Akkor az előbb definiált magfüggvényes predikciós stratégia univerzálisan konzisztens a stacionárius, ergodikus $\{(\mathbf{X}_n, Y_n)\}_{-\infty}^{\infty}$ folyamatok azon osztályára, amelyre $|Y_0| \leq B$.*

5.4. Univerzálisan konzisztens legközelebbi szomszéd stratégia

Ismét $h^{(k,\ell)}$ elemi predikciók egy kétindexes halmazát vezetjük be, ahol k jelöli a közelmúltnak azt a hosszát, amelyre illeszkedéseket keresünk, és minden ℓ -re választunk egy p_ℓ -t $(0, 1)$ -ből úgy, hogy

$$\lim_{\ell \rightarrow \infty} p_\ell = 0, \quad (5.11)$$

és legyen

$$\bar{\ell} = \lfloor p_\ell n \rfloor.$$

Az n időpontban rögzített k -ra és ℓ -re ($n > k + \bar{\ell} + 1$) az elemi prediktor megkeresi az $\mathbf{X}_{n-k}^n, Y_{n-k}^{n-1}$ szegmens $\bar{\ell}$ darab legközelebbi szomszédját, ezek lesznek az illeszkedések, és az elemi predikció az illeszkedéseket követő Y_t -k átlag. Formálisan, legyen

$$J_n^{(k,\ell)} = \left\{ k < t < n : (\mathbf{X}_{t-k}^t, Y_{t-k}^{t-1}) \text{ egyike a } (\mathbf{X}_{n-k}^n, Y_{n-k}^{n-1}) \bar{\ell} \text{ NN-jének az } (\mathbf{X}_1^{k+1}, Y_1^k), \dots, (\mathbf{X}_{n-k-1}^{n-1}, Y_{n-k-1}^{n-2}) \text{ halmazból } \right\}$$

és legyen az elemi prediktor a

$$h_n^{(k,\ell)}(\mathbf{X}_1^n, Y_1^{n-1}) = \frac{\sum_{\{t \in J_n^{(k,\ell)}\}} Y_t}{|J_n^{(k,\ell)}|}$$

lokális átlag, ha az összeg nem üres, és 0 egyébként. Végül kombináljuk az elemi prediktorokat úgy, mint korábban: (5.2), (5.3), (5.4) és (5.5).

5.4.3. tétel *Tegyük fel (5.11)-t és azt, hogy minden \mathbf{s} vektor esetén az*

$$\|(\mathbf{X}_1^{k+1}, Y_1^k) - \mathbf{s}\|$$

valószínűségi változó eloszlásfüggvénye folytonos. Akkor az előbb definiált legközelebbi szomszéd predikciós stratégia univerzálisan konzisztens a stacionárius, ergodikus $\{(\mathbf{X}_n, Y_n)\}_{-\infty}^{\infty}$ folyamatok azon osztályára, amelyre $|Y_0| \leq B$.

5.5. Univerzálisan konzisztens általánosított lineáris stratégia

Az előző szakaszokban az elemi prediktorok lokális átlagolás elvén alapultak. Ebben a szakaszban az empirikus hibaminimalizálás típusú elvet használjuk, amely először a Györfi and Lugosi [2002] cikkben jelent meg. Ismét $h^{(k,\ell)}$, $k, \ell = 1, 2, \dots$ kétindexes elemi prediktorokat vezetünk be. Legyenek $\{\phi_j^{(k)}\}_{j=1}^{\ell}$ valós értékű függvények $(\mathbb{R}^d)^{(k+1)} \times \mathbb{R}^k$ -n. A $h_n^{(k,\ell)}$ elemi prediktor becslései a következő alakúak

$$h_n^{(k,\ell)}(\mathbf{x}_1^n, y_1^{n-1}) = \sum_{j=1}^{\ell} c_{n,j} \phi_j^{(k)}(\mathbf{x}_{n-k}^n, y_{n-k}^{n-1}),$$

ahol a $c_{n,j}$ együtthatókat úgy számítjuk ki, hogy minimalizáljuk a

$$\sum_{t=k+1}^{n-1} \left(\sum_{j=1}^{\ell} c_j \phi_j^{(k)}(\mathbf{x}_{t-k}^t, y_{t-k}^{t-1}) - y_t \right)^2 \quad (5.12)$$

kritériumot, ha $n > k + 1$, egyébként az együtthatók nullák.

Az elemi prediktorokat ugyanúgy kombináljuk, mint korábban: (5.2), (5.3), (5.4) és (5.5).

5.5.4. tétel (GYÖRFI AND LUGOSI [2002]) *Tegyük fel, hogy $|\phi_j^{(k)}| \leq 1$ és minden rögzített k -ra a*

$$\left\{ \sum_{j=1}^{\ell} c_j \phi_j^{(k)}; (c_1, \dots, c_{\ell}), \ell = 1, 2, \dots \right\}$$

halmaz sűrű a $d(k+1) + k$ változós folytonos függvények terében. Akkor az előbb definiált általánosított lineáris predikciós stratégia univerzálisan konzisztens a stacionárius, ergodikus $\{(\mathbf{X}_n, Y_n)\}_{-\infty}^{\infty}$ folyamatok azon osztályára, amelyre $|Y_0| \leq B$.

6. fejezet

Alakfelismerés

6.1. Bayes döntés

Egy statisztikai következtetési problémában adott egy \mathbf{X} d dimenziós megfigyelési vektor, és \mathbf{X} alapján a statisztikus következtetést von le egy nem megfigyelhető, véges sok értéket felvevő Y valószínűségi változóról. Tegyük fel, hogy Y az értékeit a $\{1, 2, \dots, M\}$ halmazból veszi. Ha ez a statisztikai következtetés egy döntés, akkor a döntés egy

$$g : \mathbb{R}^d \rightarrow \{1, 2, \dots, M\}$$

döntésfüggvénnyel adható meg. Ha $g(\mathbf{X}) \neq Y$, akkor a döntés hibázik.

A Bayes döntési problémában bevezetünk egy $C(y, y') \geq 0$ költséget, amelyik a költség akkor, ha a címke $Y = y$ és a döntés $g(\mathbf{X}) = y'$. Egy g döntésfüggvény esetén a kockázat a költség várható értéke:

$$R(g) = \mathbb{E}\{C(Y, g(\mathbf{X}))\}.$$

A Bayes döntési problémában a cél a kockázat minimalizálása, azaz keressük azt a $g^* : \mathbb{R}^d \rightarrow \{1, 2, \dots, M\}$ döntésfüggvényt, amelyre

$$R(g^*) = \min_{g: \mathbb{R}^d \rightarrow \{1, 2, \dots, M\}} R(g). \quad (6.1)$$

A g^* -ot Bayes döntésnek hívjuk, és $R^* = R(g^*)$ jelöli a Bayes kockázatot.

Az a posteriori valószínűségekre vezessük be az alábbi jelöléseket:

$$P_y(\mathbf{X}) = \mathbb{P}\{Y = y \mid \mathbf{X}\}.$$

Definiáljuk g^* döntésfüggvényt

$$g^*(\mathbf{X}) = \arg \min_{y'} \sum_{y=1}^M C(y, y') P_y(\mathbf{X}).$$

Ha $\arg \min$ nem egyértelmű, akkor válasszuk azt a legkisebb y' -öt, amelyik minimalizálja $\sum_{y=1}^m C(y, y')P_y(\mathbf{X})$ -et. Ebből a definícióból következik, hogy tetszőleges g döntésfüggvényre

$$\sum_{y=1}^m C(y, g^*(\mathbf{X}))P_y(\mathbf{X}) \leq \sum_{y=1}^M C(y, g(\mathbf{X}))P_y(\mathbf{X}). \quad (6.2)$$

6.1.1. tétel *Tetszőleges g döntésfüggvényre*

$$R(g^*) \leq R(g).$$

BIZONYÍTÁS. Egy tetszőleges g döntésfüggvényre számoljuk ki a kockázatot!

$$\begin{aligned} R(g) &= \mathbb{E}\{C(Y, g(\mathbf{X}))\} \\ &= \mathbb{E}\{\mathbb{E}\{C(Y, g(\mathbf{X})) \mid \mathbf{X}\}\} \\ &= \mathbb{E}\left\{\sum_{y=1}^m \sum_{y'=1}^M C(y, y')\mathbb{P}\{Y = y, g(\mathbf{X}) = y' \mid \mathbf{X}\}\right\} \\ &= \mathbb{E}\left\{\sum_{y=1}^m \sum_{y'=1}^M C(y, y')\mathbb{I}_{\{g(\mathbf{X})=y'\}}\mathbb{P}\{Y = y \mid \mathbf{X}\}\right\} \\ &= \mathbb{E}\left\{\sum_{y=1}^M C(y, g(\mathbf{X}))P_y(\mathbf{X})\right\}. \end{aligned}$$

(6.2) miatt

$$\begin{aligned} R(g) &= \mathbb{E}\left\{\sum_{y=1}^M C(y, g(\mathbf{X}))P_y(\mathbf{X})\right\} \\ &\geq \mathbb{E}\left\{\sum_{y=1}^M C(y, g^*(\mathbf{X}))P_y(\mathbf{X})\right\} \\ &= R(g^*). \end{aligned}$$

□

A legfontosabb költségfüggvény az úgynevezett 0 – 1 költség:

$$C(y, y') = \begin{cases} 1 & \text{ha } y \neq y', \\ 0 & \text{ha } y = y'. \end{cases}$$

0 – 1 esetén a megfelelő kockázat a hibavalószínűség:

$$R(g) = \mathbb{E}\{C(Y, g(\mathbf{X}))\} = \mathbb{E}\{\mathbb{I}_{\{Y \neq g(\mathbf{X})\}}\} = \mathbb{P}\{Y \neq g(\mathbf{X})\},$$

és a Bayes döntés a következő alakú:

$$g^*(\mathbf{X}) = \arg \min_{y'} \sum_{y=1}^M C(y, y') P_y(\mathbf{X}) = \arg \min_{y'} \sum_{y \neq y'} P_y(\mathbf{X}) = \arg \max_{y'} P_{y'}(\mathbf{X}),$$

amelyet maximum a posteriori döntésnek is hívunk.

Ha az \mathbf{X} megfigyelési vektornak van sűrűségfüggvénye, akkor a Bayes döntésnek van egy ekvivalens formája. Legyen f az \mathbf{X} sűrűségfüggvénye, azaz

$$\mathbb{P}\{\mathbf{X} \in B\} = \int_B f(\mathbf{x}) d\mathbf{x}$$

és legyen f_y az \mathbf{X} feltételes sűrűségfüggvénye adott y mellett, azaz

$$\mathbb{P}\{\mathbf{X} \in B \mid Y = y\} = \int_B f_y(\mathbf{x}) d\mathbf{x}$$

és az a priori valószínűségek

$$q_y = \mathbb{P}\{Y = y\},$$

akkor egyszerűen belátható, hogy

$$P_y(\mathbf{X}) = \mathbb{P}\{Y = y \mid \mathbf{X} = \mathbf{x}\} = \frac{q_y f_y(\mathbf{x})}{f(\mathbf{x})},$$

ezért

$$\begin{aligned} g^*(\mathbf{x}) &= \arg \min_{y'} \sum_{y=1}^M C(y, y') P_y(\mathbf{x}) \\ &= \arg \min_{y'} \sum_{y=1}^M C(y, y') \frac{q_y f_y(\mathbf{x})}{f(\mathbf{x})} \\ &= \arg \min_{y'} \sum_{y=1}^M C(y, y') q_y f_y(\mathbf{x}). \end{aligned}$$

A 6.1.1. tétel bizonyításából következik az optimális kockázatra egy formula:

$$R(g^*) = \mathbb{E} \left\{ \min_{y'} \sum_{y=1}^M C(y, y') P_y(\mathbf{X}) \right\}.$$

Ha \mathbf{X} -nek van sűrűségfüggvénye, akkor

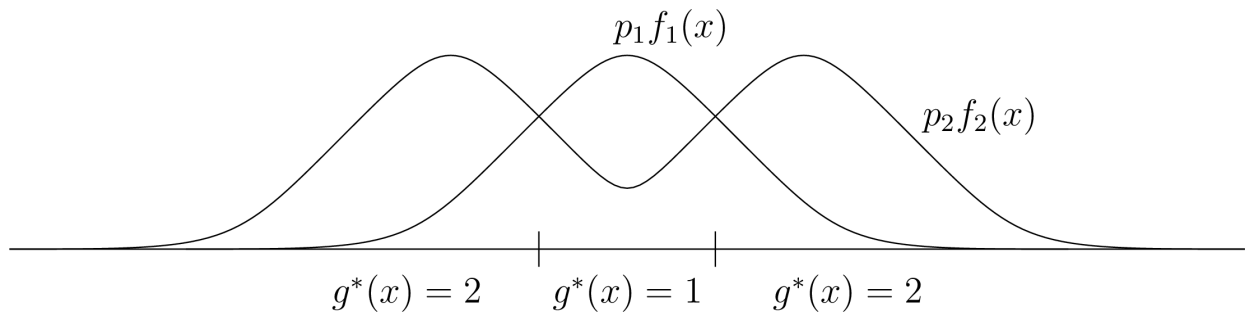
$$\begin{aligned} R(g^*) &= \mathbb{E} \left\{ \min_{y'} \sum_{y=1}^M C(y, y') \frac{q_y f_y(\mathbf{X})}{f(\mathbf{X})} \right\} \\ &= \int_{\mathbb{R}^d} \min_{y'} \sum_{y=1}^M C(y, y') \frac{q_y f_y(\mathbf{x})}{f(\mathbf{x})} f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \min_{y'} \sum_{y=1}^M C(y, y') q_y f_y(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

0 – 1 költség esetén

$$R(g^*) = \mathbb{E} \left\{ \min_{y'} (1 - P_{y'}(\mathbf{X})) \right\},$$

amely sűrűségfüggvény esetén a következő formájú:

$$R(g^*) = \int_{\mathbb{R}^d} \min_{y'} (f(\mathbf{x}) - q_{y'} f_{y'}(\mathbf{x})) d\mathbf{x} = 1 - \int_{\mathbb{R}^d} \max_{y'} q_{y'} f_{y'}(\mathbf{x}) d\mathbf{x}.$$



6.1. ábra. Bayes döntés.

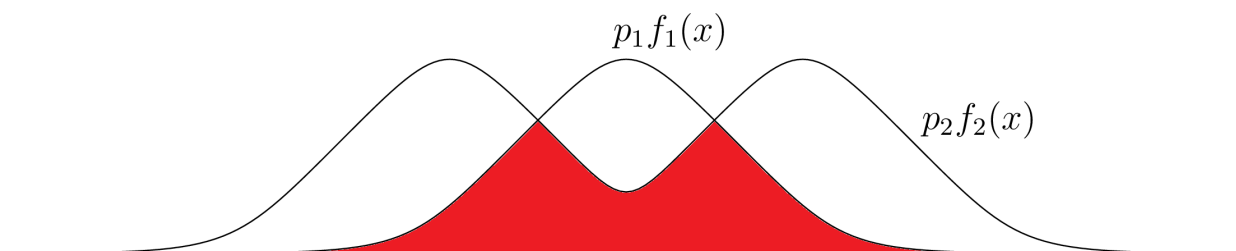
$M = 2$ esetén

$$R(g^*) = \mathbb{E} \{ \min(P_1(\mathbf{X}), P_2(\mathbf{X})) \},$$

és sűrűségfüggvények esetén

$$R(g^*) = \int_{\mathbb{R}^d} \min(q_1 f_1(\mathbf{x}), q_2 f_2(\mathbf{x})) d\mathbf{x}.$$

A 6.1 ábra illusztrálja a Bayes döntést, és a 6.2 ábrán a piros terület egyenlő a Bayes hibavalószínűséggel.



6.2. ábra. Bayes hibavalószínűség.

6.2. A Bayes döntés közelítése

Gyakorlatban a $\{P_y(\mathbf{X})\}$ aposzteriori valószínűségek ismeretlenek. Ha adott az aposzteriori valószínűségek $\{\hat{P}_y(\mathbf{X})\}$ közelítései, akkor azok alapján bevezethetjük a Bayes döntés egy közelítését:

$$\hat{g}(\mathbf{X}) = \arg \min_{y'} \sum_{y=1}^M C(y, y') \hat{P}_y(\mathbf{X}).$$

Az a kérdés, hogy milyen jól közelíti $R(\hat{g})$ az R^* -ot.

6.1 lemma Legyen $C_{max} = \max_{y, y'} C(y, y')$, akkor

$$0 \leq R(\hat{g}) - R(g^*) \leq 2C_{max} \sum_{y=1}^M \mathbb{E} \{ |P_y(\mathbf{X}) - \hat{P}_y(\mathbf{X})| \}.$$

BIZONYÍTÁS. Mivel

$$\begin{aligned}
R(\hat{g}) - R(g^*) &= \mathbb{E} \left\{ \sum_{y=1}^M C(y, \hat{g}(\mathbf{X})) P_y(\mathbf{X}) \right\} - \mathbb{E} \left\{ \sum_{y=1}^M C(y, g^*(\mathbf{X})) P_y(\mathbf{X}) \right\} \\
&= \mathbb{E} \left\{ \sum_{y=1}^M C(y, \hat{g}(\mathbf{X})) P_y(\mathbf{X}) - \sum_{y=1}^M C(y, \hat{g}(\mathbf{X})) \hat{P}_y(\mathbf{X}) \right\} \\
&\quad + \mathbb{E} \left\{ \sum_{y=1}^M C(y, \hat{g}(\mathbf{X})) \hat{P}_y(\mathbf{X}) - \sum_{y=1}^M C(y, g^*(\mathbf{X})) \hat{P}_y(\mathbf{X}) \right\} \\
&\quad + \mathbb{E} \left\{ \sum_{y=1}^M C(y, g^*(\mathbf{X})) \hat{P}_y(\mathbf{X}) - \sum_{y=1}^M C(y, g^*(\mathbf{X})) P_y(\mathbf{X}) \right\},
\end{aligned}$$

ezért \hat{g} definíciójából következik, hogy

$$\sum_{y=1}^M C(y, \hat{g}(\mathbf{X})) \hat{P}_y(\mathbf{X}) - \sum_{y=1}^M C(y, g^*(\mathbf{X})) \hat{P}_y(\mathbf{X}) \leq 0,$$

tehát

$$\begin{aligned}
R(\hat{g}) - R(g^*) &\leq \mathbb{E} \left\{ \sum_{y=1}^M C(y, \hat{g}(\mathbf{X})) |P_y(\mathbf{X}) - \hat{P}_y(\mathbf{X})| \right\} \\
&\quad + \mathbb{E} \left\{ \sum_{y=1}^M C(y, g^*(\mathbf{X})) |\hat{P}_y(\mathbf{X}) - P_y(\mathbf{X})| \right\} \\
&\leq 2C_{max} \sum_{y=1}^M \mathbb{E} \left\{ |P_y(\mathbf{X}) - \hat{P}_y(\mathbf{X})| \right\}.
\end{aligned}$$

□

A közelítőleg maximum a posteriori döntés esetén az 6.1 lemma egyenlőtlensége kicsit javítható:

$$0 \leq R(\hat{g}) - R(g^*) \leq \sum_{y=1}^M \mathbb{E} \left\{ |P_y(\mathbf{X}) - \hat{P}_y(\mathbf{X})| \right\}.$$

Az 6.1 lemma alapján hatékony alakfelismerési eljárásokat lehet konstruálni. Az aposzteriori valószínűségek egyben feltételes várható értékek is, tehát regressziós függvények

$$\mathbb{P}\{Y = y | \mathbf{X} = \mathbf{x}\} = \mathbb{E}\{\mathbb{I}_{\{Y=y\}} | \mathbf{X} = \mathbf{x}\} = m^{(y)}(\mathbf{x}).$$

Adott $D_n = \{(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)\}$ adatok esetén az $m^{(y)}$ -nak az $m_n^{(y)}$ becslését megkonstruálhatjuk a

$$D_n^{(y)} = \{(\mathbf{X}_1, \mathbb{I}_{\{Y_1=y\}}), \dots, (\mathbf{X}_n, \mathbb{I}_{\{Y_n=y\}})\}$$

adatokból, amiből levezethető egy alakfelismerési eljárás:

$$g_n(\mathbf{x}) = \arg \max_{1 \leq y \leq M} m_n^{(y)}(\mathbf{x}), \quad (6.3)$$

ami a g^* egy közelítése. Ha az $m_n^{(y)}$ becslések közel vannak $m^{(y)}$ -hoz, akkor a szóbanforgó alakfelismerési eljárás hibaválószerűsége is közel van az optimálishoz. (A részleteket lásd Devroye, Györfi, and Lugosi [1996] könyvben.)

6.3. Alakfelismerés idősorokra

Ebben a szakaszban is idősorok predikciójával foglalkozunk, de a négyzetes költség helyett 0–1 költséget alkalmazunk, tehát a predikció nem egy becslés, hanem egy döntés. Legyen az $\{(\mathbf{X}_n, Y_n)\}_{n=1}^{\infty}$ idősor pároknak egy stacionárius, ergodikus sorozata, ahol a párok $\mathbb{R}^d \times \{0, 1\}$ -ből veszik az értékeiket, tehát az Y -ok bináris értékűek. A probléma az, hogy döntsünk Y_n -ről az idősor $(\mathbf{X}_1^n, Y_1^{n-1})$ múltja alapján.

A döntési, osztályozási feladatot egy alakfelismerési stratégiával formalizáljuk, amelyik $f = \{f_t\}_{t=1}^{\infty}$ döntésfüggvények egy sorozata:

$$f_t : (\mathbb{R}^d)^t \times \{0, 1\}^{t-1} \rightarrow \{0, 1\},$$

tehát a t időpontban Y_t -re a döntésünk $f_t(\mathbf{X}_1^t, Y_1^{t-1})$. A \mathbf{X}_1^n, Y_1^n szegmens esetén az *átlagos 0–1 hibát*

$$R_n(f) = \frac{1}{n} \sum_{t=1}^n \mathbb{I}_{\{f_t(\mathbf{X}_1^t, Y_1^{t-1}) \neq Y_t\}}$$

definiálja.

Az elvi optimum ebben az esetben is ismert. Algoet [1994] bizonyította, hogy minden f alakfelismerési stratégiára és $\{(\mathbf{X}_n, Y_n)\}_{n=1}^{\infty}$, stacionárius, ergodikus idősorra

$$\liminf_{n \rightarrow \infty} R_n(f) \geq R^* \quad 1 \text{ valószínűséggel}, \quad (6.4)$$

ahol

$$R^* = \mathbb{E} \left\{ \min \left(\mathbb{P}\{Y_0 = 1 | \mathbf{X}_{-\infty}^0, Y_{-\infty}^{-1}\}, \mathbb{P}\{Y_0 = 0 | \mathbf{X}_{-\infty}^0, Y_{-\infty}^{-1}\} \right) \right\}.$$

Mivel az optimális alakfelismerési stratégia felhasználja az idősor többdimenziós eloszlásait, ezért gyakorlati szempontból is fontos kérdés, hogy vajon csak az idősor múltját megfigyelve vajon közelíthető-e az optimális stratégia.

6.1 definíció *Egy f alakfelismerési stratégiát univerzálisan konzisztensnek nevezzük, ha minden $\{\mathbf{X}_n, Y_n\}_{n=-\infty}^{\infty}$ stacionárius, ergodikus idősorra*

$$\lim_{n \rightarrow \infty} R_n(f) = R^* \quad 1 \text{ valószínűséggel.}$$

Most mutatunk univerzálisan konzisztens alakfelismerési stratégiákat. Legyen $g_t(\mathbf{X}_1^t, Y_1^{t-1})$ egyike azon predikciós eljárásoknak, amelyeket a 5.2 vagy a 5.3 vagy a 5.4 szakaszokban vezettünk be, és legyen scheme:

$$f_t(\mathbf{X}_1^t, Y_1^{t-1}) = \begin{cases} 1 & \text{ha } g_t(\mathbf{X}_1^t, Y_1^{t-1}) > 1/2 \\ 0 & \text{egyébként.} \end{cases}$$

Ennek a szakasznak a fő eredménye, hogy ez az alakfelismerési stratégia univerzálisan konzisztens:

6.3.2. tétel (GYÖRFI ÉS OTTUCSÁK [2007]) *Tegyük fel, hogy a 5.2.1. vagy a 5.3.2. vagy a 5.4.3. tétel feltételei teljesülnek. Akkor az előbb definiált f alakfelismerési stratégia univerzálisan konzisztens, azaz*

$$\lim_{n \rightarrow \infty} R_n(f) = R^* \quad 1 \text{ valószínűséggel}$$

minden $\{(\mathbf{X}_n, Y_n)\}_{n=-\infty}^{\infty}$ stacionárius, ergodikus idősorra.

A 6.3.2. tétel bizonyításához felhasználjuk a következőt:

6.1 következmény *A 5.2.1. vagy a 5.3.2. vagy a 5.4.3. tétel feltételei esetén*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (\mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\} - g_i(\mathbf{X}_1^i, Y_1^{i-1}))^2 = 0 \quad 1 \text{ valószínűséggel.}$$

BIZONYÍTÁS. A 5.2.1. vagy a 5.3.2. vagy a 5.4.3. tétel miatt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (Y_i - g_i(\mathbf{X}_1^i, Y_1^{i-1}))^2 = L^* \quad 1 \text{ valószínűséggel.}$$

Tekintsük a következő felbontást:

$$\begin{aligned} & (Y_i - g_i(\mathbf{X}_1^i, Y_1^{i-1}))^2 \\ = & (Y_i - \mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\})^2 \\ & + 2(Y_i - \mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\}) (\mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\} - g_i(\mathbf{X}_1^i, Y_1^{i-1})) \\ & + (\mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\} - g_i(\mathbf{X}_1^i, Y_1^{i-1}))^2. \end{aligned}$$

Az ergodtétel miatt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\})^2 = L^* \quad 1 \text{ valószínűséggel.}$$

Elég azt megmutatni, hogy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\}) (\mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\} - g_i(\mathbf{X}_1^i, Y_1^{i-1})) = 0 \quad (6.5)$$

1 valószínűséggel, ami viszont következik Chow [1965]-nak a martingáldifferenciákra vonatkozó nagy számok erős törvényéből (lásd Stout [1974, Theorem 3.3.1]). Ez azt állítja, hogy ha $\{Z_i\}$ egy martingáldifferencia úgy, hogy

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}Z_n^2}{n^2} < \infty, \quad (6.6)$$

akkor

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Z_i = 0 \quad 1 \text{ valószínűséggel.}$$

Ekkor (6.5) következik a Chow törvényből, mivel a

$$Z_i = (Y_i - \mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\}) (\mathbb{E}\{Y_i | \mathbf{X}_{-\infty}^i, Y_{-\infty}^{i-1}\} - g_i(\mathbf{X}_1^i, Y_1^{i-1}))$$

martingáldifferenciák korlátosak $4B^2$ korláttal. □

A 6.3.2. TÉTEL BIZONYÍTÁSA. A (6.4) miatt elég azt megmutatni, hogy

$$\limsup_{n \rightarrow \infty} R_n(f) \leq R^* \quad 1 \text{ valószínűséggel.}$$

A 6.1 következményből adódik, hogy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n (\mathbb{E}\{Y_t \mid \mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}\} - g_t(\mathbf{X}_1^t, Y_1^{t-1}))^2 = 0 \quad 1 \text{ valószínűséggel.} \quad (6.7)$$

Vezessük be a teljes végtelen múltat felhasználó Bayes osztályozót:

$$f_t^*(\mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}) = \begin{cases} 1 & \text{ha } \mathbb{P}\{Y_t = 1 \mid \mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}\} > 1/2 \\ 0 & \text{egyébként,} \end{cases}$$

és az ő átlagos 0 – 1 hibáját :

$$R_n(f^*) = \frac{1}{n} \sum_{t=1}^n \mathbb{I}_{\{f_t^*(\mathbf{x}_{-\infty}^t, Y_{-\infty}^{t-1}) \neq Y_t\}}.$$

Legyen

$$\bar{R}_n(f) = \frac{1}{n} \sum_{t=1}^n \mathbb{P}\{f_t(\mathbf{X}_1^t, Y_1^{t-1}) \neq Y_t \mid \mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}\}$$

és

$$\bar{R}_n(f^*) = \frac{1}{n} \sum_{t=1}^n \mathbb{P}\{f_t^*(\mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}) \neq Y_t \mid \mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}\}.$$

Akkor

$$R_n(f) - \bar{R}_n(f) \rightarrow 0 \quad 1 \text{ valószínűséggel}$$

és

$$R_n(f^*) - \bar{R}_n(f^*) \rightarrow 0 \quad 1 \text{ valószínűséggel,}$$

mivel azok korlátos martingáldifferenciák átlagai. Továbbá az ergodtételeből következik, hogy

$$\bar{R}_n(f^*) \rightarrow R^* \quad 1 \text{ valószínűséggel,}$$

ezért csak azt kell megmutatni, hogy

$$\limsup_{n \rightarrow \infty} (\bar{R}_n(f) - \bar{R}_n(f^*)) \leq 0 \quad 1 \text{ valószínűséggel.}$$

A 6.1 lemma miatt

$$\begin{aligned}
\bar{R}_n(f) - \bar{R}_n(f^*) &= \frac{1}{n} \sum_{t=1}^n \left(\mathbb{P}\{f_t(\mathbf{X}_1^t, Y_1^{t-1}) \neq Y_t \mid \mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}\} \right. \\
&\quad \left. - \mathbb{P}\{f_t^*(\mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}) \neq Y_t \mid \mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}\} \right) \\
&\leq 2 \frac{1}{n} \sum_{t=1}^n |\mathbb{E}\{Y_t \mid \mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}\} - g_t(\mathbf{X}_1^t, Y_1^{t-1})| \\
&\leq 2 \sqrt{\frac{1}{n} \sum_{t=1}^n |\mathbb{E}\{Y_t \mid \mathbf{X}_{-\infty}^t, Y_{-\infty}^{t-1}\} - g_t(\mathbf{X}_1^t, Y_1^{t-1})|^2} \\
&\rightarrow 0 \quad 1 \text{ valószínűséggel,}
\end{aligned}$$

ahol az utolsó lépésben (6.7)-t használtuk. □

7. fejezet

Sűrűségfüggvénybecslés

7.1. Miért becsüljük sűrűségfüggvényt: az L_1 hiba

A klasszikus nemparaméteres példa az

$$F(\mathbf{x}) = \mathbb{P}\{\mathbf{X} < \mathbf{x}\}$$

eloszlásfüggvény becslése független, azonos eloszlású $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ adatokból, amelyek \mathbb{R}^d -ből veszik értékeiket. Itt egyrészt az

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathbf{x}_i < \mathbf{x}\}}$$

empirikus eloszlásfüggvénynek a konstrukciója eloszlásfüggetlen (univerzális), másrészt az alapvető konzisztencia, a Glivenko-Cantelli tétel és minden eloszlásfüggvényre érvényes. A Glivenko-Cantelli tétel szerint minden F -re

$$\lim_{n \rightarrow \infty} \sup_{\mathbf{x} \in \mathbb{R}^d} |F_n(\mathbf{x}) - F(\mathbf{x})| = 0$$

1 valószínűséggel.

A Glivenko-Cantelli tétel valóban eloszlásfüggetlen, és a Kolmogorov-Szmirnov távolsággal értelmezett konvergencia egyenletes konvergenciát jelent, ezért úgy tűnik, hogy nem is kell tovább vizsgálni. Ugyanakkor, ha például az empirikus eloszlásfüggvény alapján akarunk megoldani egy döntési problémát, akkor az használhatatlan, amennyiben a szóbanforgó eloszlásfüggvény folytonos. Kiderül, hogy a Kolmogorov-Szmirnov

távolságnál szigorúbb hibakritérium kell, amelyik nemcsak d dimenziós szögletekre hasonlítja össze az empirikus és az igazi eloszlást. Természetszerűleg adódik a variációs távolság bevezetése: ha μ és ν az \mathbb{R}^d -n két eloszlás, akkor a μ és ν *variációs távolságát* a

$$V(\mu, \nu) = \sup_A |\mu(A) - \nu(A)|$$

szupremummal definiáljuk, ahol a szupremumot minden A Borel halmazra vesszük. A Scheffé tétel alább azt mondja, hogy a variációs távolság a megfelelő sűrűségfüggvények L_1 távolságának a fele.

7.1.1. tétel (SCHEFFÉ [1947]) *Ha μ és ν abszolút folytonos f és g sűrűségfüggvényekkel, akkor*

$$\int_{\mathbb{R}^d} |f(\mathbf{x}) - g(\mathbf{x})| d\mathbf{x} = 2V(\mu, \nu).$$

(A

$$L_1(f, g) = \int_{\mathbb{R}^d} |f(\mathbf{x}) - g(\mathbf{x})| d\mathbf{x} \tag{7.1}$$

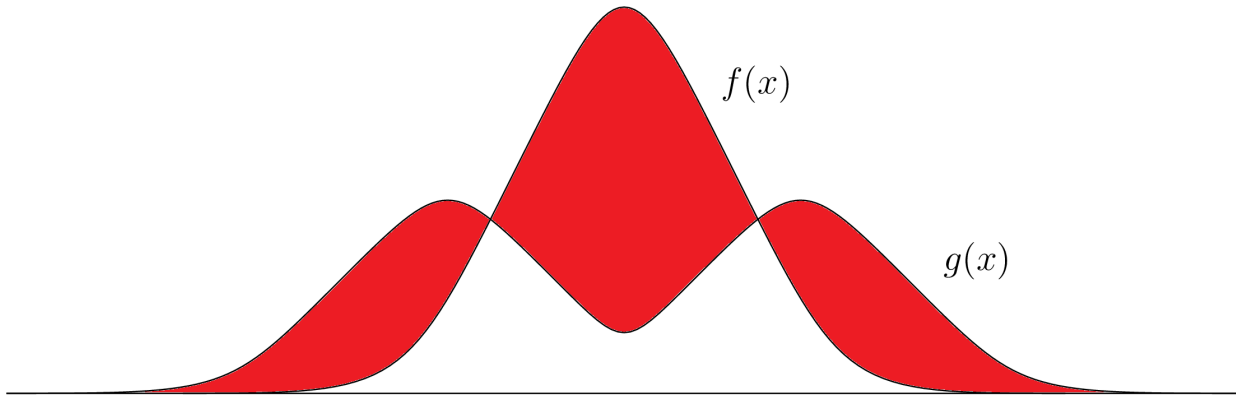
mennyiséget L_1 távolságnak hívjuk.)

BIZONYÍTÁS. Vegyük észre, hogy

$$\begin{aligned} V(\mu, \nu) &= \sup_A |\mu(A) - \nu(A)| \\ &= \sup_A \left| \int_A f - \int_A g \right| \\ &= \sup_A \left| \int_A (f - g) \right| \\ &= \int_{f>g} (f - g) \\ &= \int_{g>f} (g - f) \\ &= \frac{1}{2} \int |f - g|. \end{aligned}$$

□

A 7.1 ábrán a piros terület egyenlő az f és g sűrűségfüggvények L_1 távolságával.



7.1. ábra. L_1 távolság.

Megjegyezzük, hogy a Scheffé tételből következik a varációs távolság egy ekvivalens definíciója:

$$V(\mu, \nu) = \frac{1}{2} \sup_{\{A_j\}} \sum_j |\mu(A_j) - \nu(A_j)|, \quad (7.2)$$

ahol szupremumot az összes véges, Borel mérhető $\{A_j\}$ partícióra vesszük.

A \mathbf{X} eloszlására vezessük be a

$$\mu(A) = \mathbb{P}\{\mathbf{X} \in A\}$$

jelölést. A következőkben tegyük fel, hogy a μ eloszlásnak van sűrűségfüggvénye, amelyet f -fel jelölünk:

$$\mu(A) = \int_A f(\mathbf{x}) d\mathbf{x}.$$

A $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ független, azonos eloszlású adatokból becsülhetjük az f sűrűségfüggvényt, és a becslést $f_n(\mathbf{x}) = f_n(\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_n)$ -nel jelöljük. Egy ilyen sűrűségfüggvénybecslésből természetes módon vezethetünk le egy μ_n^* eloszlásbecslést:

$$\mu_n^*(A) = \int_A f_n(\mathbf{x}) d\mathbf{x}.$$

Akkor a Scheffé tétel miatt

$$V(\mu, \mu_n^*) = \frac{1}{2} \int_{\mathbb{R}^d} |f(\mathbf{x}) - f_n(\mathbf{x})| d\mathbf{x},$$

tehát ha az f_n sűrűségfüggvénybecslés L_1 konzisztens, azaz

$$\lim_{n \rightarrow \infty} \int |f(\mathbf{x}) - f_n(\mathbf{x})| \, d\mathbf{x} = 0$$

1 valószínűséggel, akkor a megfelelő μ_n^* eloszlásbecslés konzisztens variációs távolságban:

$$\lim_{n \rightarrow \infty} V(\mu, \mu_n^*) = 0$$

1 valószínűséggel.

7.2. A hisztogram

Jelölje μ_n a

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathbf{x}_i \in A\}}$$

az empirikus eloszlást. Legyen $\mathcal{P}_n = \{A_{n1}, A_{n2}, \dots\}$ az \mathbb{R}^d egy olyan partíciója, amelyre a A_{nj} celláknak pozitív és véges a λ térfogata (Lebesgue mértéke). Akkor a hisztogram definíciója

$$f_n(\mathbf{x}) = \frac{\mu_n(A_n(\mathbf{x}))}{\lambda(A_n(\mathbf{x}))},$$

ahol

$$A_n(\mathbf{x}) = A_{nj}, \text{ ha } \mathbf{x} \in A_{nj}.$$

A \mathcal{P}_n partícióra a leggyakoribb példa a kockás partíció, amikor a cellák h_n oldalhosszúságú d dimenziós kockák. Ebben a speciális esetben

$$f_n(\mathbf{x}) = \frac{\mu_n(A_n(\mathbf{x}))}{h_n^d}.$$

A következő tétel a hisztogram L_1 konzisztenciáját mondja ki tetszőleges sűrűségfüggvényre.

7.2.2. tétel *Tegyük fel, hogy minden origó közepű S gömbre teljesül, hogy*

$$\lim_{n \rightarrow \infty} \sup_{j: A_{nj} \cap S \neq \emptyset} \text{diam}(A_{nj}) = 0$$

és

$$\lim_{n \rightarrow \infty} \frac{|\{j : A_{nj} \cap S \neq \emptyset\}|}{n} = 0,$$

akkor

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \int |f(\mathbf{x}) - f_n(\mathbf{x})| \, d\mathbf{x} \right\} = 0$$

1 valószínűséggel.

BIZONYÍTÁS. A háromszögegyenlőtlenség miatt

$$\int |f_n(\mathbf{x}) - f(\mathbf{x})| \, d\mathbf{x} \leq \underbrace{\int |f_n(\mathbf{x}) - \mathbb{E}f_n(\mathbf{x})| \, d\mathbf{x}}_{\text{variációs tag}} + \underbrace{\int |\mathbb{E}f_n(\mathbf{x}) - f(\mathbf{x})| \, d\mathbf{x}}_{\text{torzítás}}.$$

A hisztogram cellánként konstans, ezért

$$\int |f_n(\mathbf{x}) - \mathbb{E}f_n(\mathbf{x})| \, d\mathbf{x} = \sum_j \int_{A_{nj}} |f_n(\mathbf{x}) - \mathbb{E}f_n(\mathbf{x})| \, d\mathbf{x} = \sum_j |\mu_n(A_{nj}) - \mu(A_{nj})|.$$

Legyen $M_n = |\{j : A_{nj} \cap S \neq \emptyset\}|$, és válasszuk meg a cellák számozását úgy, hogy $A_{nj} \cap S \neq \emptyset$, $j = 1, \dots, M_n$. A tétel feltételei miatt

$$\frac{M_n}{n} \rightarrow 0.$$

Legyen

$$S_n = \bigcup_{j=1}^{M_n} A_{nj}.$$

Akkor

$$\int |f_n(\mathbf{x}) - \mathbb{E}f_n(\mathbf{x})| \, d\mathbf{x} \leq \sum_{j=1}^{M_n} |\mu_n(A_{nj}) - \mu(A_{nj})| + \mu_n(S_n^c) + \mu(S_n^c),$$

ezért a Cauchy-Schwarz és a Jensen egyenlőtlenségből következik, hogy

$$\begin{aligned}
\mathbb{E} \left\{ \int |f_n(\mathbf{x}) - \mathbb{E}f_n(\mathbf{x})| d\mathbf{x} \right\} &\leq \sum_{j=1}^{M_n} \mathbb{E}\{|\mu_n(A_{nj}) - \mu(A_{nj})|\} + 2\mu(S_n^c) \\
&\leq \sum_{j=1}^{M_n} \sqrt{\mathbb{E}\{|\mu_n(A_{nj}) - \mu(A_{nj})|^2\}} + 2\mu(S^c) \\
&\leq \sum_{j=1}^{M_n} \sqrt{\frac{\mu(A_{nj})}{n}} + 2\mu(S^c) \\
&\leq \sqrt{\frac{M_n}{n}} + 2\mu(S^c) \\
&\rightarrow 2\mu(S^c).
\end{aligned} \tag{7.3}$$

Az S gömb tetszőleges, tehát

$$\mathbb{E} \left\{ \int |f_n(\mathbf{x}) - \mathbb{E}f_n(\mathbf{x})| d\mathbf{x} \right\} \rightarrow 0.$$

A torzítás esetén

$$\mathbb{E}f_n(\mathbf{x}) = \frac{\mu(A_n(\mathbf{x}))}{\lambda(A_n(\mathbf{x}))} = \frac{1}{\lambda(A_n(\mathbf{x}))} \int_{A_n(\mathbf{x})} f(\mathbf{z}) d\mathbf{z} = \int f(\mathbf{z})K_n(\mathbf{x}, \mathbf{z}) d\mathbf{z},$$

ahol

$$K_n(\mathbf{x}, \mathbf{z}) = \frac{\mathbb{I}_{\{\mathbf{z} \in A_n(\mathbf{x})\}}}{\lambda(A_n(\mathbf{x}))}.$$

Ekkor

$$\int |\mathbb{E}f_n(\mathbf{x}) - f(\mathbf{x})| d\mathbf{x} = \int \left| \int f(\mathbf{z})K_n(\mathbf{x}, \mathbf{z}) d\mathbf{z} - f(\mathbf{x}) \right| d\mathbf{x}.$$

Ha f folytonos és egy kompakt halmazon kívül nulla, akkor egyenletesen is folytonos, és így az

$$\int |\mathbb{E}f_n(\mathbf{x}) - f(\mathbf{x})| d\mathbf{x} \leq \int \int |f(\mathbf{z}) - f(\mathbf{x})|K_n(\mathbf{x}, \mathbf{z}) d\mathbf{z}d\mathbf{x} \tag{7.4}$$

egyenlőtlenségből következik, hogy

$$\int |\mathbb{E}f_n(\mathbf{x}) - f(\mathbf{x})| d\mathbf{x} \rightarrow 0.$$

A folytonos függvények halmaza sűrű L_1 -ben, ezért tetszőleges f sűrűségfüggvényhez és $\varepsilon > 0$ -hoz létezik egy \tilde{f} sűrűségfüggvény úgy, hogy az folytonos és egy kompakt halmazon kívül nulla, és

$$\int |f(\mathbf{x}) - \tilde{f}(\mathbf{x})| \, d\mathbf{x} < \varepsilon.$$

Ekkor

$$\begin{aligned} & \int |f(\mathbf{x}) - \mathbb{E}f_n(\mathbf{x})| \, d\mathbf{x} \\ = & \int \left| f(\mathbf{x}) - \int f(\mathbf{z})K_n(\mathbf{x}, \mathbf{z}) \, d\mathbf{z} \right| \, d\mathbf{x} \\ \leq & \int |f(\mathbf{x}) - \tilde{f}(\mathbf{x})| \, d\mathbf{x} + \int \left| \tilde{f}(\mathbf{x}) - \int \tilde{f}(\mathbf{z})K_n(\mathbf{x}, \mathbf{z}) \, d\mathbf{z} \right| \, d\mathbf{x} \\ & + \int \left| \int \tilde{f}(\mathbf{z})K_n(\mathbf{x}, \mathbf{z}) \, d\mathbf{z} - \int f(\mathbf{z})K_n(\mathbf{x}, \mathbf{z}) \, d\mathbf{z} \right| \, d\mathbf{x} \\ \leq & \varepsilon + \int \left| \tilde{f}(\mathbf{x}) - \int \tilde{f}(\mathbf{z})K_n(\mathbf{x}, \mathbf{z}) \, d\mathbf{z} \right| \, d\mathbf{x} \\ & + \int \left(\int |\tilde{f}(\mathbf{z}) - f(\mathbf{z})|K_n(\mathbf{x}, \mathbf{z}) \, d\mathbf{x} \right) \, d\mathbf{z} \\ = & \varepsilon + \int \left| \tilde{f}(\mathbf{x}) - \int \tilde{f}(\mathbf{z})K_n(\mathbf{x}, \mathbf{z}) \, d\mathbf{z} \right| \, d\mathbf{x} + \int |\tilde{f}(\mathbf{z}) - f(\mathbf{z})| \, d\mathbf{z} \\ \rightarrow & 2\varepsilon. \end{aligned}$$

□

A következő tétel a hisztogram L_1 hibájának a konvergenciasebességét mutatja be, amennyiben a sűrűségfüggvényre teljesülnek bizonyos feltételek.

7.2.3. tétel *Tegyük fel, hogy f Lipschitz folytonos, azaz*

$$|f(\mathbf{x}) - f(\mathbf{z})| \leq C\|\mathbf{x} - \mathbf{z}\|,$$

és egy S gömbön kívül nulla. Ha a \mathcal{P}_n partíció kockás, és a cellák h_n oldalhosszúsága, akkor a f_n hisztogramra

$$\mathbb{E} \int |f - f_n| \leq \frac{c_1}{\sqrt{nh_n^d}} + c_2h_n,$$

tehát a

$$h_n = c_3n^{-\frac{1}{d+2}}$$

választásra

$$\mathbb{E} \int |f - f_n| \leq c_4 n^{-\frac{1}{d+2}}.$$

BIZONYÍTÁS. A variációs tagra (7.3)-ból következik, hogy

$$\mathbb{E} \left\{ \int |f_n(\mathbf{x}) - \mathbb{E}f_n(\mathbf{x})| d\mathbf{x} \right\} \leq \sqrt{\frac{M_n}{n}} \leq \sqrt{\frac{\lambda(S)}{nh_n^d}}.$$

A torzítás esetén (7.4) miatt azt kapjuk, hogy

$$\begin{aligned} \int |\mathbb{E}f_n(\mathbf{x}) - f(\mathbf{x})| d\mathbf{x} &\leq \int \int |f(\mathbf{z}) - f(\mathbf{x})| K_n(\mathbf{x}, \mathbf{z}) d\mathbf{z} d\mathbf{x} \\ &\leq \int \int C \|\mathbf{z} - \mathbf{x}\| K_n(\mathbf{x}, \mathbf{z}) d\mathbf{z} d\mathbf{x} \\ &\leq \int \int Ch_n \sqrt{d} K_n(\mathbf{x}, \mathbf{z}) d\mathbf{z} d\mathbf{x} \\ &\leq Ch_n \sqrt{d} \lambda(S). \end{aligned}$$

□

7.3. Magfüggvényes sűrűségfüggvénybecslés

Válasszunk egy $K(\mathbf{x})$ sűrűségfüggvényt, amit magfüggvénynek hívunk. Egy pozitív h_n sávszélesség esetén a magfüggvényes sűrűségfüggvénybecslést a

$$f_n(\mathbf{x}) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right)$$

képlettel definiáljuk.

7.3.4. tétel *Ha*

$$\lim_{n \rightarrow \infty} h_n = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} nh_n^d = \infty.$$

akkor az f_n magfüggvényes sűrűségfüggvénybecslésre

$$\lim_{n \rightarrow \infty} \mathbb{E} \int |f(\mathbf{x}) - f_n(\mathbf{x})| d\mathbf{x} = 0.$$

Példák magfüggvényre:

- naiv vagy ablak magfüggvény

$$K(\mathbf{x}) = c\mathbb{I}_{\{\|\mathbf{x}\|\leq 1\}},$$

- Gauss magfüggvény

$$K(\mathbf{x}) = ce^{-\|\mathbf{x}\|^2}.$$

- Cauchy magfüggvény

$$K(\mathbf{x}) = \frac{c}{1 + \|\mathbf{x}\|^{d+1}}.$$

- Epanechnikov magfüggvény

$$K(\mathbf{x}) = c(1 - \|\mathbf{x}\|^2)\mathbb{I}_{\{\|\mathbf{x}\|\leq 1\}}.$$

7.3.5. tétel Tegyük fel, hogy f differenciálható Lipschitz folytonos gradienssel, és nulla egy S gömbön kívül. Akkor az f_n magfüggvényes sűrűségfüggvénybecslésre

$$\mathbb{E} \int |f - f_n| \leq \frac{c_1}{\sqrt{nh_n^d}} + c_2 h_n^2,$$

tehát a

$$h_n = c_3 n^{-\frac{1}{d+4}}$$

választásra

$$\mathbb{E} \int |f - f_n| \leq c_4 n^{-\frac{2}{d+4}}.$$

Az L_1 sűrűségfüggvénybecslésről olvasnivalónak ajánljuk a Devroye, Györfi [1985], Devroye [1987] és Devroye, Lugosi [2001] könyveket.

8. fejezet

Egyszerű hipotézisek vizsgálata

8.1. α szintű tesztek

Ebben a fejezetben olyan döntési problémákat tekintünk, amikor a különböző hibázásoknak nagyon eltérőek a következményei. Ha például egy diagnosztikai feladatban $Y = 1$ azt jelenti, hogy a páciens beteg, míg $Y = 0$ azt, hogy egészséges, akkor $Y = 0$ esetén a hibás döntés az, hogy a páciens beteg, és akkor a hibás döntésnek csak a felesleges orvosi kezelés a következménye. $Y = 1$ esetén viszont a hibázás azt jelenti, hogy a páciens egészséges, és ekkor az elmaradt orvosi kezelés miatt rosszabbodhat a páciens állapota.

Az $Y = 0$ eseményt nullhipotézisnek hívjuk, és \mathcal{H}_0 -al jelöljük, míg az $Y = 1$ eseményt alternatív hipotézisnek nevezzük, és \mathcal{H}_1 -gyel jelöljük. A döntést, más szóval tesztet egy úgynevezett $A \subset \mathbb{R}^d$ elfogadási tartománnyal adjuk meg, ugyanis elfogadjuk a \mathcal{H}_0 nullhipotézist, ha $\mathbf{X} \in A$, egyébként elutasítjuk \mathcal{H}_0 -t, azaz elfogadjuk a \mathcal{H}_1 alternatív hipotézist. A A^c halmazt kritikus tartománynak hívjuk.

Legyen P_0 illetve P_1 az \mathbf{X} megfigyelési vektor eloszlása \mathcal{H}_0 illetve \mathcal{H}_1 esetén. Két hibavalószínűség lehetséges:

- Elsőfajú hiba, amikor \mathcal{H}_0 nullhipotézis esetén elutasítjuk \mathcal{H}_0 -t. Ennek a hibának az értéke $P_0(A^c)$.
- Másodfajú hiba, amikor a \mathcal{H}_1 alternatív hipotézis esetén elutasítjuk \mathcal{H}_1 -et. Ennek a hibának az értéke $P_1(A)$.

Természetesen, csökkenthetjük $P_0(A^c)$ elsőfajú hibát a $P_1(A)$ másodfajú hiba rovására. Ebben az esetben az optimalizálási feladatot úgy tűzzük ki, hogy minimalizáljuk a másodfajú hibát azzal a megszorítással, hogy az elsőfajú hiba legfeljebb $0 < \alpha < 1$:

$$\min_{A: P_0(A^c) \leq \alpha} P_1(A). \quad (8.1)$$

Ennek az optimalizálási feladatnak a megoldásában fontos szerepet játszik a Neyman-Pearson lemma.

8.1.1. tétel (NEYMAN, PEARSON [1933]) *Tegyük fel, hogy a P_0 illetve a P_1 eloszlásoknak van sűrűségfüggvényük, amelyeket f_0 -al illetve f_1 -gyel jelölünk:*

$$P_0(B) = \int_B f_0(\mathbf{x})d\mathbf{x} \quad \text{és} \quad P_1(B) = \int_B f_1(\mathbf{x})d\mathbf{x}.$$

Egy $\gamma > 0$ esetén legyen

$$A_\gamma = \{\mathbf{x} : f_0(\mathbf{x}) \geq \gamma f_1(\mathbf{x})\}$$

egy elfogadási tartomány. Ha egy A elfogadási tartományra

$$P_0(A^c) \leq P_0(A_\gamma),$$

akkor

$$P_1(A) \geq P_1(A_\gamma).$$

BIZONYÍTÁS. A tétel feltételeiből egyenlőtlenségeknek a következő láncolata vezethető le:

$$\begin{aligned} P_0(A^c) &\leq P_0(A_\gamma^c) \\ P_0(A^c \cap A_\gamma) + P_0(A^c \cap A_\gamma^c) &\leq P_0(A \cap A_\gamma) + P_0(A^c \cap A_\gamma^c) \\ \int_{A^c \cap A_\gamma} f_0(x)dx &\leq \int_{A \cap A_\gamma^c} f_0(x)dx. \end{aligned}$$

Az A_γ definíciója miatt

$$\gamma \int_{A^c \cap A_\gamma} f_1(\mathbf{x})d\mathbf{x} \leq \int_{A^c \cap A_\gamma} f_0(\mathbf{x})d\mathbf{x} \leq \int_{A \cap A_\gamma^c} f_0(\mathbf{x})d\mathbf{x} \leq \gamma \int_{A \cap A_\gamma^c} f_1(\mathbf{x})d\mathbf{x},$$

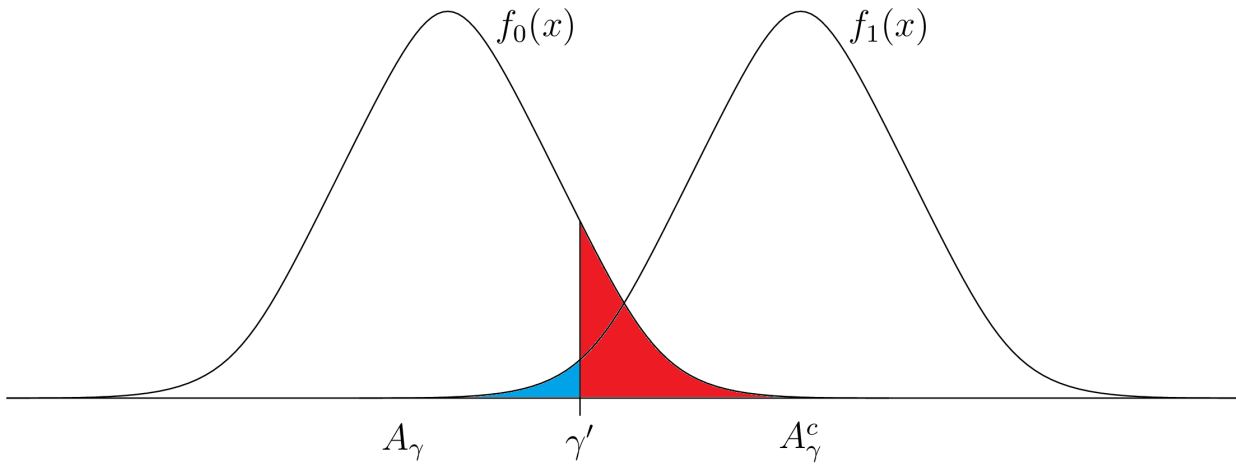
ezért az előző láncolat megfordításából azt kapjuk, hogy

$$P_1(A^c) \leq P_1(A_\gamma^c).$$

□

A 8.1 ábrán a kék terület illusztrálja az elsőfajú hibát, míg a piros a másodfajút. Ha egy $0 < \alpha < 1$ -re létezik $\gamma = \gamma(\alpha)$ úgy, hogy az megoldja a

$$P_0(A_\gamma^c) = \alpha$$



8.1. ábra. Elsőfajú és másodfajú hiba.

egyenletet, akkor a Neyman-Pearson lemmából következik, hogy a (8.1) megoldásához elég a A_γ alakú elfogadási tartományokat tekinteni, azaz

$$\min_{A: P_0(A^c) \leq \alpha} P_1(A) = \min_{A_\gamma: P_0(A_\gamma^c) \leq \alpha} P_1(A_\gamma).$$

Ekkor A_γ -t *legerősebb α szintű tesztnek* hívjuk.

A Neyman-Pearson lemma miatt bevezetjük a

$$T(\mathbf{X}) = \frac{f_0(\mathbf{X})}{f_1(\mathbf{X})}$$

likelihood-hányadost, és így elfogadjuk a \mathcal{H}_0 nullhipotézist, ha $T(\mathbf{X}) \geq \gamma$.

1. PÉLDA. A Neyman-Pearson lemma illusztrációjaként tekintsük azt a kísérleti példát, amikor a nullhipotézis esetén az \mathbf{X} komponensei független, normális eloszlású valószínűségi változók $m = m_0 > 0$ várható értékkel és σ^2 varianciával, míg az alternatív hipotézis esetén az \mathbf{X} komponensei független, normális eloszlású valószínűségi változók $m_1 = 0$ várható értékkel és ugyanazzal a σ^2 varianciával. Akkor

$$f_0(\mathbf{x}) = f_0(x_1, \dots, x_d) = \prod_{i=1}^d \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - m)^2}{2\sigma^2}} \right)$$

és

$$f_1(\mathbf{x}) = f_1(x_1, \dots, x_d) = \prod_{i=1}^d \left(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x_i^2}{2\sigma^2}} \right).$$

Az

$$\frac{f_0(\mathbf{X})}{f_1(\mathbf{X})} \geq \gamma$$

elfogadás azt jelenti, hogy

$$-\sum_{i=1}^d \frac{(X_i - m)^2}{2\sigma^2} + \sum_{i=1}^d \frac{X_i^2}{2\sigma^2} \geq \ln \gamma,$$

vagy ezzel ekvivalensen

$$\sum_{i=1}^d (2X_i m - m^2) \geq 2\sigma^2 \ln \gamma.$$

Ez teszt tehát akkor fogadja el a nullhipotézist, ha

$$\frac{1}{d} \sum_{i=1}^d X_i \geq \frac{2\sigma^2 \ln \gamma / d + m^2}{2m} = \frac{\sigma^2 \ln \gamma}{dm} + \frac{m}{2} =: \gamma'.$$

A kapott teszt a $\sum_{i=1}^d X_i / d$ lineáris statisztika alapján dönt, ezért problémaként csak azon γ' kritikus érték meghatározása marad, amelyre a teszt α szintű, azaz az alsófajú hiba éppen α :

$$\mathbb{P}_0 \left\{ \frac{1}{d} \sum_{i=1}^d X_i \leq \gamma' \right\} = \alpha.$$

A nullhipotézis esetén $\frac{1}{d} \sum_{i=1}^d X_i$ eloszlása normális m várható értékkel és σ^2/d szórásnégyzettel, ezért

$$\mathbb{P}_0 \left\{ \frac{1}{d} \sum_{i=1}^d X_i \leq \gamma' \right\} = \Phi \left(\frac{\gamma' - m}{\sigma/\sqrt{d}} \right),$$

ahol Φ jelöli a standard normális eloszlásfüggvényt, és így az α szintű teszt γ' kritikus értéke megoldja a

$$\Phi \left(-\frac{m - \gamma'}{\sigma/\sqrt{d}} \right) = \alpha$$

egyenletet, azaz

$$\gamma' = m - \Phi^{-1}(1 - \alpha)\sigma/\sqrt{d}.$$

1. MEGJEGYZÉS. Számos gyakorlati feladatban ugyan az \mathbf{X} nem normális eloszlású, de a d olyan nagy, hogy hivatkozhatunk a centrális határeloszlástételre, és így a

$$\ln \frac{f_0(\mathbf{X})}{f_1(\mathbf{X})}$$

log-likelihood hányados eloszlása közelítőleg normális. Ezek után az 1. példa gondolatmenetét kiterjesztjük úgy, hogy \mathcal{H}_0 esetén a log-likelihood hányados eloszlása közelítőleg normális m_0 várható értékkel és σ_0^2 szórásnégyzettel. Ha a tesztet úgy definiáljuk, hogy elfogadjuk \mathcal{H}_0 -t, amennyiben

$$\ln \frac{f_0(\mathbf{X})}{f_1(\mathbf{X})} \geq \gamma',$$

ahol

$$\gamma' = m_0 - \Phi^{-1}(1 - \alpha)\sigma_0,$$

akkor ez a teszt közelítőleg α szintű teszt.

8.2. ϕ divergenciák

Az első- és másodfajú hibák együttes csökkentése úgy lehetséges, ha egyetlen megfigyelési vektor helyett adott megfigyelési vektorok egy sorozata. Ezt a lehetőséget hívjuk ismételt megfigyelésnek. Az ismételt megfigyelések analízisében fontos szerepet játszanak az információs divergenciák. Csiszár Imre [1967] vezette be a ϕ divergenciát. Legyen $\phi : (0, \infty) \rightarrow \mathbb{R}$ egy konvex függvény, amelyet a folytonosság megtartásával kiterjesztünk $[0, \infty)$ -re, és $\phi(1) = 0$. A μ és ν valószínűségeloszlásra legyen λ egy σ -véges domináló mérték, például $\lambda = \mu + \nu$. Vezessük be az

$$f = \frac{d\mu}{d\lambda}$$

és a

$$g = \frac{d\nu}{d\lambda}$$

jelöléseket. Akkor a μ és a ν ϕ divergenciáját a

$$D_\phi(\mu, \nu) = \int_{\mathbb{R}^d} \phi\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) g(\mathbf{x}) \lambda(d\mathbf{x}) \quad (8.2)$$

képlettel definiáljuk.

A Jensen egyenlőtlenségből következik a ϕ divergencia legfontosabb tulajdonsága:

$$D_\phi(\mu, \nu) = \int_{\mathbb{R}^d} \phi\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) g(\mathbf{x}) \lambda(d\mathbf{x}) \geq \phi\left(\int_{\mathbb{R}^d} \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) \lambda(d\mathbf{x})\right) = \phi(1) = 0.$$

Ez azt jelenti, hogy $D_\phi(\mu, \nu) \geq 0$, és ha $\mu = \nu$, akkor $D_\phi(\mu, \nu) = 0$. Ha még ráadásul ϕ szigorúan konvex 1-ben, akkor $D_\phi(\mu, \nu) = 0$ akkor és csak akkor, ha $\mu = \nu$.

Mutatunk néhány példát:

- A

$$\phi_1(t) = |t - 1|$$

esetén kapjuk az L_1 távolságot:

$$D_{\phi_1}(\mu, \nu) = \int_{\mathbb{R}^d} |f(\mathbf{x}) - g(\mathbf{x})| \lambda(d\mathbf{x}).$$

- A

$$\phi_2(t) = (\sqrt{t} - 1)^2$$

választásnál a divergenciát *négyzetes Hellinger távolságnak* hívjuk:

$$\begin{aligned} D_{\phi_2}(\mu, \nu) &= \int_{\mathbb{R}^d} \left(\sqrt{f(\mathbf{x})} - \sqrt{g(\mathbf{x})}\right)^2 \lambda(d\mathbf{x}) \\ &= 2 \left(1 - \int_{\mathbb{R}^d} \sqrt{f(\mathbf{x})g(\mathbf{x})} \lambda(d\mathbf{x})\right). \end{aligned}$$

- A

$$\phi_3(t) = -\ln t,$$

esetén jutunk az I divergenciához, más néven relatív entrópiához vagy Kullback-Leibler divergenciához:

$$I(\mu, \nu) = D_{\phi_3}(\mu, \nu) = \int_{\mathbb{R}^d} \ln\left(\frac{g(\mathbf{x})}{f(\mathbf{x})}\right) g(\mathbf{x}) \lambda(d\mathbf{x}).$$

-

$$\phi_4(t) = (t - 1)^2,$$

esetén kapjuk a χ^2 divergenciát:

$$\chi^2(\mu, \nu) = D_{\phi_4}(\mu, \nu) = \int_{\mathbb{R}^d} \frac{(f(\mathbf{x}) - g(\mathbf{x}))^2}{g(\mathbf{x})} \lambda(d\mathbf{x}).$$

Belátható, hogy ϕ divergencia egy ekvivalens definíciójához jutunk úgy, hogy

$$D_\phi(\mu, \nu) = \sup_{\mathcal{P}} \sum_j \phi \left(\frac{\mu(A_j)}{\nu(A_j)} \right) \nu(A_j), \quad (8.3)$$

ahol a szupremumot az \mathbb{R}^d összes véges, Borel mérhető $\mathcal{P} = \{A_j\}$ partíciójára vesszük.

Ennek az ekvivalenciának az egyik fő oka az, hogy a Jensen egyenlőtlenség miatt minden $\mathcal{P} = \{A_j\}$ partícióra

$$\begin{aligned} D_\phi(\mu, \nu) &= \int_{\mathbb{R}^d} \phi \left(\frac{f(\mathbf{x})}{g(\mathbf{x})} \right) g(\mathbf{x}) \lambda(d\mathbf{x}) \\ &= \sum_j \int_{A_j} \phi \left(\frac{f(\mathbf{x})}{g(\mathbf{x})} \right) g(\mathbf{x}) \lambda(d\mathbf{x}) \\ &= \sum_j \frac{1}{\nu(A_j)} \int_{A_j} \phi \left(\frac{f(\mathbf{x})}{g(\mathbf{x})} \right) g(\mathbf{x}) \lambda(d\mathbf{x}) \nu(A_j) \\ &\geq \sum_j \phi \left(\frac{1}{\nu(A_j)} \int_{A_j} \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) \lambda(d\mathbf{x}) \right) \nu(A_j) \\ &= \sum_j \phi \left(\frac{\mu(A_j)}{\nu(A_j)} \right) \nu(A_j). \end{aligned} \quad (8.4)$$

Ha a $\mathcal{P}_1, \mathcal{P}_2, \dots$ partíciók sorozata finomodó, azaz minden $A \in \mathcal{P}_{n+1}$ cella részhalmaza egy $A' \in \mathcal{P}_n$ -nak, akkor megmutatjuk, hogy

$$\sum_{A \in \mathcal{P}_n} \phi \left(\frac{\mu(A)}{\nu(A)} \right) \nu(A) \uparrow.$$

Ez a tulajdonság ismét a Jensen egyenlőtlenség következménye:

$$\begin{aligned}
\sum_{A' \in \mathcal{P}_{n+1}} \phi \left(\frac{\mu(A')}{\nu(A')} \right) \nu(A') &= \sum_{A \in \mathcal{P}_n} \left(\sum_{A' \in \mathcal{P}_{n+1}, A' \subset A} \phi \left(\frac{\mu(A')}{\nu(A')} \right) \nu(A') \right) \\
&= \sum_{A \in \mathcal{P}_n} \left(\sum_{A' \in \mathcal{P}_{n+1}, A' \subset A} \phi \left(\frac{\mu(A')}{\nu(A')} \right) \frac{\nu(A')}{\nu(A)} \right) \nu(A) \\
&\geq \sum_{A \in \mathcal{P}_n} \phi \left(\sum_{A' \in \mathcal{P}_{n+1}, A' \subset A} \frac{\mu(A')}{\nu(A')} \frac{\nu(A')}{\nu(A)} \right) \nu(A) \\
&= \sum_{A \in \mathcal{P}_n} \phi \left(\frac{\mu(A)}{\nu(A)} \right) \nu(A).
\end{aligned}$$

Ebből következik, hogy finimodó $\mathcal{P}_1, \mathcal{P}_2, \dots$ partíciók esetén

$$\sum_{A \in \mathcal{P}_n} \phi \left(\frac{\mu(A)}{\nu(A)} \right) \nu(A) \uparrow \sup_{\mathcal{P}_n} \sum_{A \in \mathcal{P}_n} \phi \left(\frac{\mu(A)}{\nu(A)} \right) \nu(A).$$

A $\mathcal{P}_1, \mathcal{P}_2, \dots$ partíciók egy sorozatát aszimptotikusan finomnak nevezzük, ha minden origó közepű S gömbre

$$\lim_{n \rightarrow \infty} \max_{A \in \mathcal{P}_n, A \cap S \neq \emptyset} \text{diam}(A) = 0. \quad (8.5)$$

Megmutatható, hogy ha a $\mathcal{P}_1, \mathcal{P}_2, \dots$ partíciók sorozata aszimptotikusan finom, akkor

$$\sum_{A \in \mathcal{P}_n} \phi \left(\frac{\mu(A)}{\nu(A)} \right) \nu(A) \uparrow \int_{\mathbb{R}^d} \phi \left(\frac{f(\mathbf{x})}{g(\mathbf{x})} \right) g(\mathbf{x}) \lambda(d\mathbf{x}).$$

Ezt az utolsó lépést ellenőrizzük az L_1 távolság speciális esetében. (Lásd (9.7)-et.) Az általános esetben bevezethetjük $\frac{f(\mathbf{x})}{g(\mathbf{x})}$ cellánként konstans közelítését:

$$F_n(\mathbf{x}) := \frac{\mu(A)}{\nu(A)} \text{ ha } \mathbf{x} \in A.$$

Akkor

$$\sum_{A \in \mathcal{P}_n} \phi \left(\frac{\mu(A)}{\nu(A)} \right) \nu(A) = \int_{\mathbb{R}^d} \phi(F_n(\mathbf{x})) g(\mathbf{x}) \lambda(d\mathbf{x})$$

és

$$F_n(\mathbf{x}) \rightarrow \frac{f(\mathbf{x})}{g(\mathbf{x})}$$

majdnem minde \mathbf{x} mod λ -ra, ahol $g(\mathbf{x}) > 0$ úgy, hogy

$$\int_{\mathbb{R}^d} \phi(F_n(\mathbf{x})) g(\mathbf{x}) \lambda(d\mathbf{x}) \rightarrow \int_{\mathbb{R}^d} \phi\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) g(\mathbf{x}) \lambda(d\mathbf{x}).$$

8.3. Ismételt megfigyelések

Mint már az előzőekben említettük, az első- és másodfajú hibák együttes csökkentése úgy lehetséges, ha egyetlen \mathbf{X} megfigyelési vektor helyett adott megfigyelési vektorok egy n hosszú $\mathbf{X}_1, \dots, \mathbf{X}_n$ egy sorozata úgy, hogy \mathcal{H}_0 esetén $\mathbf{X}_1, \dots, \mathbf{X}_n$ független és P_0 eloszlású vektorok, míg \mathcal{H}_1 esetén $\mathbf{X}_1, \dots, \mathbf{X}_n$ független és P_1 eloszlású vektorok. Ekkor a likelihood hányados a következő alakú:

$$T(\mathbf{X}) = \frac{f_0(\mathbf{X}_1) \cdot \dots \cdot f_0(\mathbf{X}_n)}{f_1(\mathbf{X}_1) \cdot \dots \cdot f_1(\mathbf{X}_n)}.$$

Az alább ismerttetett Stein lemma szerint létezik olyan teszt, hogy mind az elsőfajú hiba α_n , mind a másodfajú hiba β_n nullához tart.

A Stein lemmához emlékeztetünk az *I divergencia* definíciójára:

$$I(P_0, P_1) = D(f_0, f_1) = \int_{\mathbb{R}^d} f_0(\mathbf{x}) \ln \frac{f_0(\mathbf{x})}{f_1(\mathbf{x})} d\mathbf{x}. \quad (8.6)$$

8.3.2. tétel (LÁSD CHERNOFF [1952]) *Minden $0 < \delta < D(f_0, f_1)$ -re létezik egy teszt úgy, hogy az elsőfajú hiba*

$$\alpha_n \rightarrow 0,$$

és a másodfajú hibára

$$\beta_n \leq e^{-n(D(f_0, f_1) - \delta)} \rightarrow 0.$$

BIZONYÍTÁS. Legyen a teszt olyan, hogy elfogadja a \mathcal{H}_0 nullhipotézist, ha

$$\frac{f_0(\mathbf{X}_1) \cdot \dots \cdot f_0(\mathbf{X}_n)}{f_1(\mathbf{X}_1) \cdot \dots \cdot f_1(\mathbf{X}_n)} \geq e^{n(D(f_0, f_1) - \delta)},$$

vagy ekvivalensen

$$\frac{1}{n} \sum_{i=1}^n \ln \frac{f_0(\mathbf{X}_i)}{f_1(\mathbf{X}_i)} \geq D(f_0, f_1) - \delta.$$

\mathcal{H}_0 esetén a nagy számok erős törvénye miatt

$$\frac{1}{n} \sum_{i=1}^n \ln \frac{f_0(\mathbf{X}_i)}{f_1(\mathbf{X}_i)} \rightarrow D(f_0, f_1)$$

1 valószínűséggel, ezért az α_n elsőfajú hibára

$$\alpha_n = \mathbb{P}_0 \left\{ \frac{1}{n} \sum_{i=1}^n \ln \frac{f_0(\mathbf{X}_i)}{f_1(\mathbf{X}_i)} < D(f_0, f_1) - \delta \right\} \rightarrow 0.$$

A β_n másodfajú hibával kapcsolatban a következő egyszerű korlátot kapjuk:

$$\begin{aligned} & \beta_n \\ &= \mathbb{P}_1 \left\{ \frac{f_0(\mathbf{X}_1) \cdot \dots \cdot f_0(\mathbf{X}_n)}{f_1(\mathbf{X}_1) \cdot \dots \cdot f_1(\mathbf{X}_n)} \geq e^{n(D(f_0, f_1) - \delta)} \right\} \\ &= \int_{\left\{ \frac{f_0(\mathbf{x}_1) \cdot \dots \cdot f_0(\mathbf{x}_n)}{f_1(\mathbf{x}_1) \cdot \dots \cdot f_1(\mathbf{x}_n)} \geq e^{n(D(f_0, f_1) - \delta)} \right\}} f_1(\mathbf{x}_1) \cdot \dots \cdot f_1(\mathbf{x}_n) d\mathbf{x}_1, \dots, d\mathbf{x}_n \\ &\leq e^{-n(D(f_0, f_1) - \delta)} \int_{\left\{ \frac{f_0(\mathbf{x}_1) \cdot \dots \cdot f_0(\mathbf{x}_n)}{f_1(\mathbf{x}_1) \cdot \dots \cdot f_1(\mathbf{x}_n)} \geq e^{n(D(f_0, f_1) - \delta)} \right\}} f_0(\mathbf{x}_1) \cdot \dots \cdot f_0(\mathbf{x}_n) d\mathbf{x}_1, \dots, d\mathbf{x}_n \\ &\leq e^{-n(D(f_0, f_1) - \delta)}. \end{aligned}$$

□

A Stein lemma bizonyításában a kritikus érték felhasználta a $D(f_0, f_1)$ divergencia ismeretét. A $D(f_0, f_1)$ ismerete nélkül az alábbi Chernoff lemma garantálja, hogy mindkét hiba nullához tart exponenciálisan gyorsan.

8.3.3. tétel (CHERNOFF [1952]). *Tekintsük azt a tesztet, amelyik elfogadja a \mathcal{H}_0 nullhipotézist, ha*

$$\frac{f_0(\mathbf{X}_1) \cdot \dots \cdot f_0(\mathbf{X}_n)}{f_1(\mathbf{X}_1) \cdot \dots \cdot f_1(\mathbf{X}_n)} \geq 1,$$

vagy ekvivalensen

$$\sum_{i=1}^n \ln \frac{f_0(\mathbf{X}_i)}{f_1(\mathbf{X}_i)} \geq 0.$$

(Ezt a tesztet maximum likelihood tesztnek nevezzük.) Akkor

$$\alpha_n \leq \left(\inf_{s>0} \int_{\mathbb{R}^d} f_1(\mathbf{x})^s f_0(\mathbf{x})^{1-s} d\mathbf{x} \right)^n$$

és

$$\beta_n \leq \left(\inf_{s>0} \int_{\mathbb{R}^d} f_0(\mathbf{x})^s f_1(\mathbf{x})^{1-s} d\mathbf{x} \right)^n .$$

BIZONYÍTÁS. Alkalmazzuk a Chernoff technikát, amelyik szerint egy $s > 0$ a Markov egyenlőtlenséget használjuk:

$$\begin{aligned} \alpha_n &= \mathbb{P}_0 \left\{ \sum_{i=1}^n \ln \frac{f_0(\mathbf{X}_i)}{f_1(\mathbf{X}_i)} < 0 \right\} \\ &= \mathbb{P}_0 \left\{ s \sum_{i=1}^n \ln \frac{f_1(\mathbf{X}_i)}{f_0(\mathbf{X}_i)} > 0 \right\} \\ &= \mathbb{P}_0 \left\{ e^{s \sum_{i=1}^n \ln \frac{f_1(\mathbf{X}_i)}{f_0(\mathbf{X}_i)}} > 1 \right\} \\ &\leq \mathbb{E}_0 \left\{ e^{s \sum_{i=1}^n \ln \frac{f_1(\mathbf{X}_i)}{f_0(\mathbf{X}_i)}} \right\} \\ &= \mathbb{E}_0 \left\{ \prod_{i=1}^n \left(\frac{f_1(\mathbf{X}_i)}{f_0(\mathbf{X}_i)} \right)^s \right\} . \end{aligned}$$

\mathcal{H}_0 esetén $\mathbf{X}_1, \dots, \mathbf{X}_n$ függetlenek, ezért

$$\begin{aligned} \alpha_n &\leq \mathbb{E}_0 \left\{ \prod_{i=1}^n \left(\frac{f_1(\mathbf{X}_i)}{f_0(\mathbf{X}_i)} \right)^s \right\} \\ &= \prod_{i=1}^n \mathbb{E}_0 \left\{ \left(\frac{f_1(\mathbf{X}_i)}{f_0(\mathbf{X}_i)} \right)^s \right\} \\ &= \mathbb{E}_0 \left\{ \left(\frac{f_1(\mathbf{X}_1)}{f_0(\mathbf{X}_1)} \right)^s \right\}^n \\ &= \left(\int_{\mathbb{R}^d} \left(\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} \right)^s f_0(\mathbf{x}) d\mathbf{x} \right)^n . \end{aligned}$$

$s > 0$ tetszőleges, tehát a tétel első felét beláttuk, és a második fele hasonlóan történhet. \square

2. MEGJEGYZÉS. A Chernoff lemmából akkor következik exponenciális konvergenciasebesség, ha

$$\inf_{s>0} \int_{\mathbb{R}^d} f_1(\mathbf{x})^s f_0(\mathbf{x})^{1-s} d\mathbf{x} < 1$$

és

$$\inf_{s>0} \int_{\mathbb{R}^d} f_0(\mathbf{x})^s f_1(\mathbf{x})^{1-s} d\mathbf{x} < 1.$$

A Cauchy-Schwartz egyenlőtlenségből következik, hogy

$$\begin{aligned} \inf_{s>0} \int_{\mathbb{R}^d} f_1(\mathbf{x})^s f_0(\mathbf{x})^{1-s} d\mathbf{x} &\leq \int_{\mathbb{R}^d} f_1(\mathbf{x})^{1/2} f_0(\mathbf{x})^{1/2} d\mathbf{x} \\ &\leq \sqrt{\int_{\mathbb{R}^d} f_1(\mathbf{x}) d\mathbf{x} \int_{\mathbb{R}^d} f_0(\mathbf{x}) d\mathbf{x}} \\ &= 1, \end{aligned}$$

és egyenlőség akkor és csak akkor, ha $f_0 = f_1$. Megmutatható továbbá, hogy

$$g(s) := \int_{\mathbb{R}^d} f_1(\mathbf{x})^s f_0(\mathbf{x})^{1-s} d\mathbf{x}$$

függvény konvex, és $g(0) = 1$ és $g(1) = 1$, ezért

$$\inf_{s>0} \int_{\mathbb{R}^d} f_1(\mathbf{x})^s f_0(\mathbf{x})^{1-s} d\mathbf{x} = \inf_{1>s>0} \int_{\mathbb{R}^d} f_1(\mathbf{x})^s f_0(\mathbf{x})^{1-s} d\mathbf{x}.$$

A

$$He(f_0, f_1) = \int_{\mathbb{R}^d} f_1(\mathbf{x})^{1/2} f_0(\mathbf{x})^{1/2} d\mathbf{x} \quad (8.7)$$

mennyiséget *Hellinger integrálnak* hívjuk. Az előző levezetés szerint

$$\alpha_n \leq He(f_0, f_1)^n$$

és

$$\beta_n \leq He(f_0, f_1)^n.$$

A $D_{\phi_2}(\mu, \nu)$ négyzetes Hellinger távolságra megmutatható, hogy

$$D_{\phi_2}(\mu, \nu) = 2(1 - He(f_0, f_1)).$$

3. MEGJEGYZÉS. Az α szintű teszt fogalmán túl ily módon egy új konzisztenciához jutunk, amit *erős konzisztenciának* hívunk. Ez azt jelenti, hogy mind \mathcal{H}_0 , mind \mathcal{H}_1 esetén a teszt nem hibázik egy véletlen mintanagyság után 1 valószínűséggel. Más szóval \mathbb{P}_0 -lal

illetve \mathbb{P}_1 -gyel jelölve az eloszlást a null- illetve alternatív hipotézis esetén, a Stein lemma tesztjére azt kapjuk, hogy

$$\mathbb{P}_0\{\text{elutasítva } \mathcal{H}_0\text{-t csak véges sokszor } n\} = 1 \quad (8.8)$$

és

$$\mathbb{P}_1\{\text{elfogadva } \mathcal{H}_0\text{-t csak véges sokszor } n\} = 1. \quad (8.9)$$

Ezek az erős állítások az exponenciális Chernoff korlátból és a Borel-Cantelli lemmából következnek.

9. fejezet

Hipotézisvizsgálat egyszerű null- és összetett alternatív hipotézis esetén

9.1. A variációs távolság és az I-divergencia

Emlékeztetnénk arra, hogy az \mathbb{R}^d -n értelmezett μ és ν eloszlások *variációs távolságát* a

$$V(\mu, \nu) = \sup_A |\mu(A) - \nu(A)|,$$

szupremummal definiáltuk, ahol a szupremumot az összes A Borel halmazra vesszük. A Scheffé tétel (7.1.1. tétel) miatt a variációs távolság a megfelelő sűrűségfüggvények L_1 távolságának a fele.

A következő egyenlőtlenség, amit Pinsker egyenlőtlenségnek hívnak, a variációs távolságra ad felső becslést az I-divergencia segítségével:

9.1.1. tétel (CSISZÁR [1967], KULLBACK [1967] ÉS KEMPERMAN [1969])

$$2\{V(\mu, \nu)\}^2 \leq I(\mu, \nu). \quad (9.1)$$

BIZONYÍTÁS. Használjuk a Scheffé tétel bizonyításának a jelölését, miszerint legyen

$$A^* = \{f > g\},$$

akkor a Scheffé tétel miatt

$$V(\mu, \nu) = \mu(A^*) - \nu(A^*).$$

Továbbá (8.4)-ből következik, hogy

$$I(\mu, \nu) \geq \mu(A^*) \ln \frac{\mu(A^*)}{\nu(A^*)} + (1 - \mu(A^*)) \ln \frac{1 - \mu(A^*)}{1 - \nu(A^*)}$$

Vezessük be a

$$q = \nu(A^*) \text{ and } p = \mu(A^*) > q,$$

és a

$$h_p(q) = p \ln \frac{p}{q} + (1 - p) \ln \frac{1 - p}{1 - q}.$$

jelöléseket, akkor azt kell belátni, hogy

$$2(p - q)^2 \leq h_p(q),$$

amely következik a deriváltból:

$$\begin{aligned} \frac{d}{dq}(h_p(q) - 2(p - q)^2) &= -\frac{p}{q} + \frac{1 - p}{1 - q} + 4(p - q) \\ &= -\frac{p - q}{q(1 - q)} + 4(p - q) \\ &\leq 0. \end{aligned}$$

□

9.2. Az L_1 távolság nagy eltérése.

Legyenek adottak \mathbb{R}^d értékű $\mathbf{X}_1, \dots, \mathbf{X}_n$ véletlen vektorok, amelyek független, azonos eloszlásúak, és a közös eloszlást jelölje ν . Egy rögzített (hipotetikus) μ eloszlás esetén tekintsük a következő hipotézisvizsgálati problémát:

$$\mathcal{H}_0 : \nu = \mu \text{ szemben a } \mathcal{H}_1 : \nu \neq \mu.$$

Itt \mathcal{H}_0 egy egyszerű hipotézis, míg \mathcal{H}_1 egy összetett hipotézis.

Erre a problémára Györfi és van der Meulen [1990] vezetett be egy teszt statisztikát:

$$L_n = \sum_{j=1}^{m_n} |\mu_n(A_{n,j}) - \mu(A_{n,j})|,$$

ahol μ_n jelöli az $\mathbf{X}_1, \dots, \mathbf{X}_n$ minta esetén az empirikus eloszlást:

$$\mu_n(A) = \frac{\#\{i : \mathbf{X}_i \in A, i = 1, \dots, n\}}{n}$$

és $\mathcal{P}_n = \{A_{n,1}, \dots, A_{n,m_n}\}$ az \mathbb{R}^d egy véges partíciója.

A következő tétel megadja az L_n pontos nagy eltérés típusú jellemzését:

9.2.2. tétel (BEIRLANT, DEVROYE, GYÖRFI ÉS VAJDA [2001]). *Tegyük fel, hogy*

$$\lim_{n \rightarrow \infty} \max_j \mu(A_{n,j}) = 0 \quad (9.2)$$

és

$$\lim_{n \rightarrow \infty} \frac{m_n \ln n}{n} = 0. \quad (9.3)$$

Akkor minden $0 < \epsilon < 2$ -ra

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{L_n > \epsilon\} = -g_L(\epsilon), \quad (9.4)$$

ahol

$$g_L(\epsilon) = \inf_{0 < p < 1 - \epsilon/2} \left(p \ln \frac{p}{p + \epsilon/2} + (1 - p) \ln \frac{1 - p}{1 - p - \epsilon/2} \right). \quad (9.5)$$

Biau és Györfi [2005] mutatott egy alternatív levezetést $g_L(\epsilon)$ -ra és egy nem aszimptotikus felső korlátot.

9.2.3. tétel (BIAU ÉS GYÖRFI [2005]). *Minden $\epsilon > 0$ -ra*

$$\mathbb{P}\{L_n > \epsilon\} \leq 2^{m_n} e^{-n\epsilon^2/2}.$$

BIZONYÍTÁS. A Scheffé tétel miatt

$$L_n = \sum_{A \in \mathcal{P}_n} |\mu_n(A) - \mu(A)| = 2 \max_{A \in \sigma(\mathcal{P}_n)} (\mu_n(A) - \mu(A)),$$

ahol a $\sigma(\mathcal{P}_n)$ halmazosztály a \mathcal{P}_n partíció celláinak összes uniójából áll. Minden $s > 0$ -ra a Markov egyenlőtlenségből következik, hogy

$$\mathbb{P}\{L_n > \epsilon\} = \mathbb{P}\{L_n/2 > \epsilon/2\} = \mathbb{P}\{e^{nsL_n/2} > e^{n\epsilon/2}\} \leq \frac{\mathbb{E}\{e^{nsL_n/2}\}}{e^{n\epsilon/2}}.$$

Továbbá

$$\begin{aligned}
\mathbb{E}\{e^{snL_n/2}\} &= \mathbb{E}\left\{\max_{A \in \sigma(\mathcal{P}_n)} e^{sn(\mu_n(A) - \mu(A))}\right\} \\
&\leq \sum_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{sn(\mu_n(A) - \mu(A))}\} \\
&\leq 2^{m_n} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{sn(\mu_n(A) - \mu(A))}\} \\
&= 2^{m_n} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{sn\mu_n(A)}\} e^{-sn\mu(A)}.
\end{aligned}$$

Minden rögzített A Borel halmazra

$$\mathbb{E}\{e^{sn\mu_n(A)}\} = \mathbb{E}\{e^{s \sum_{i=1}^n \mathbb{1}_{\mathbf{x}_i \in A}}\} = \prod_{i=1}^n \mathbb{E}\{e^{s \mathbb{1}_{\mathbf{x}_i \in A}}\} = (e^s \mu(A) + 1 - \mu(A))^n.$$

Ekkor minden $s > 0$ -ra

$$\mathbb{P}\{L_n > \epsilon\} \leq 2^{m_n} \left[\max_{A \in \sigma(\mathcal{P}_n)} e^{-s(\mu(A) + \epsilon/2)} (e^s \mu(A) + 1 - \mu(A)) \right]^n.$$

Rögzített A esetén válasszuk s -et úgy, hogy

$$e^s = \frac{\mu(A) + \epsilon/2}{1 - (\mu(A) + \epsilon/2)} \frac{1 - \mu(A)}{\mu(A)},$$

akkor erre az s -re

$$\begin{aligned}
e^{-s(\mu(A) + \epsilon/2)} (e^s \mu(A) + 1 - \mu(A)) &= e^{-I((\mu(A) + \epsilon/2, 1 - \mu(A) - \epsilon/2), (\mu(A), 1 - \mu(A)))} \\
&\leq e^{-\epsilon^2/2},
\end{aligned}$$

ahol az utolsó lépésben a Pinsker egyenlőtlenséget alkalmaztuk. Ekkor

$$\mathbb{P}\{L_n > \epsilon\} \leq 2^{m_n} e^{-n\epsilon^2/2}.$$

□

1. MEGJEGYZÉS Az előző levezetés egy speciális eseteként kapjuk a Chernoff egyenlőtlenséget:

$$\mathbb{P}\{\mu_n(A) - \mu(A) \geq \epsilon\} \leq e^{-nI((\mu(A) + \epsilon/2, 1 - \mu(A) - \epsilon/2), (\mu(A), 1 - \mu(A)))}$$

és a Hoeffding [1963] egyenlőtlenséget:

$$\mathbb{P}\{\mu_n(A) - \mu(A) \geq \epsilon\} \leq e^{-2n\epsilon^2}. \quad (9.6)$$

A Hoeffding egyenlőtlenség teljesül nem feltétlenül bináris értékű változókra. Legyenek X_1, \dots, X_n független, valós értékű valószínűségi változók, tegyük fel, hogy $a, b \in \mathbb{R}$, $a < b$ számokkal $X_i \in [a, b]$. Akkor minden $\epsilon > 0$ -ra

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^n(X_i - \mathbb{E}\{X_i\})\right| > \epsilon\right\} \leq 2e^{-\frac{2n\epsilon^2}{|b-a|^2}}.$$

Ennek további finomítása a Bernstein [1946] egyenlőtlenség, amelyik figyelembe veszi a szórást is. Legyenek X_1, \dots, X_n független, valós értékű valószínűségi változók, tegyük fel, hogy $a, b \in \mathbb{R}$, $a < b$ számokkal $X_i \in [a, b]$, és legyen

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^n \text{Var}\{X_i\} > 0.$$

Akkor minden $\epsilon > 0$ -ra

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^n(X_i - \mathbb{E}\{X_i\})\right| > \epsilon\right\} \leq 2e^{-\frac{n\epsilon^2}{2\sigma^2 + 2\epsilon(b-a)/3}}.$$

9.3. L_1 távolság alapú erősen konzisztens teszt

A 9.2.3. tétel alapján bevezetünk egy erősen konzisztens tesztet, amelyik elutasítja \mathcal{H}_0 nullhipotézist, amennyiben

$$L_n > c_1 \sqrt{\frac{m_n}{n}},$$

ahol

$$c_1 > \sqrt{2 \ln 2} \approx 1.177.$$

Akkor a $\mathcal{H}_0 = \{\nu = \mu\}$ nullhipotézis esetén 9.2.3. tételből következik egy nem aszimptotikus korlát az elsőfajú hibára:

$$\mathbb{P}\left\{L_n > c_1 \sqrt{\frac{m_n}{n}}\right\} \leq 2^{m_n} e^{-nc_1^2 m_n / (2n)} = e^{-m_n(c_1^2/2 - \ln 2)} \rightarrow 0$$

Ha $m_n/\ln n \rightarrow \infty$, akkor

$$\sum_{n=1}^{\infty} \mathbb{P} \left\{ L_n > c_1 \sqrt{\frac{m_n}{n}} \right\} < \infty,$$

tehát a Borel-Cantelli miatt a $\mathcal{H}_0 = \{\nu = \mu\}$ nullhipotézis esetén ez a teszt erősen konzisztens függetlenül attól, hogy mi a μ .

Tegyük még fel, hogy a $\mathcal{P}_1, \mathcal{P}_2, \dots$ partíciók sorozata aszimptotikusan finom (lásd (8.5)). Akkor a $\mathcal{H}_1 = \{\nu \neq \mu\}$ alternatív hipotézis esetén a háromszögeyenlőtlenség miatt

$$\begin{aligned} L_n &= \sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \mu(A_{nj})| \\ &\geq \sum_{j=1}^{m_n} |\mu(A_{nj}) - \nu(A_{nj})| - \sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \nu(A_{nj})|. \end{aligned}$$

A fentiek alapján

$$\sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \nu(A_{nj})| \rightarrow 0,$$

1 valószínűséggel, míg a (8.5) feltétel és $\{\nu \neq \mu\}$ miatt

$$\sum_{j=1}^{m_n} |\mu(A_{nj}) - \nu(A_{nj})| \rightarrow 2 \sup_B |\mu(B) - \nu(B)| = 2V(\mu, \nu) > 0. \quad (9.7)$$

ezért

$$\liminf_{n \rightarrow \infty} L_n \geq 2V(\mu, \nu) > 0 \quad (9.8)$$

1 valószínűséggel, tehát $L_n > c_1 \sqrt{m_n/n}$ 1 valószínűséggel elég nagy n -re, és így megmutattuk, hogy az L_n alapú teszt erősen konzisztens a \mathcal{H}_1 alternatív hipotézis esetén is.

Be kell még látni (9.7)-t. A Barron, Györfi és van der Meulen [1992] cikk alapján válasszunk λ mértéket, amelyik dominálja μ -t és ν -t, például legyen $\lambda = \mu + \nu$. Jelölje

f a $\mu - \nu$ -nek λ szerinti Radon-Nikodym deriváltját. Akkor egyrészt

$$\begin{aligned}
 \sum_{A \in \mathcal{P}_n} |\mu(A) - \nu(A)| &= \sum_{A \in \mathcal{P}_n} \left| \int_A f \, d\lambda \right| \\
 &\leq \sum_{A \in \mathcal{P}_n} \int_A |f| \, d\lambda \\
 &= \int |f| \, d\lambda \\
 &= 2 \sup_B |\mu(B) - \nu(B)|.
 \end{aligned}$$

Másrészt egyenletesen folytonos f -re (8.5) miatt azt kapjuk, hogy

$$\sum_{A \in \mathcal{P}_n} \left| \int_A f \, d\lambda \right| \rightarrow \int |f| \, d\lambda.$$

Ha f tetszőleges, akkor minden $\delta > 0$ -ra választhatunk egy egyenletesen folytonos \tilde{f} függvényt, amelyre

$$\int |f - \tilde{f}| \, d\lambda < \delta.$$

Ekkor

$$\begin{aligned}
 \sum_{A \in \mathcal{P}_n} \left| \int_A f \, d\lambda \right| &\geq \sum_{A \in \mathcal{P}_n} \left| \int_A \tilde{f} \, d\lambda \right| - \sum_{A \in \mathcal{P}_n} \left| \int_A (f - \tilde{f}) \, d\lambda \right| \\
 &\geq \sum_{A \in \mathcal{P}_n} \left| \int_A \tilde{f} \, d\lambda \right| - \int |f - \tilde{f}| \, d\lambda \\
 &\geq \sum_{A \in \mathcal{P}_n} \left| \int_A \tilde{f} \, d\lambda \right| - \delta \\
 &\rightarrow \int |\tilde{f}| \, d\lambda - \delta \\
 &\geq \int |f| \, d\lambda - 2\delta \\
 &= 2 \sup_B |\mu(B) - \nu(B)| - 2\delta.
 \end{aligned}$$

Mivel δ tetszőleges, ezért (9.7)-t beláttuk.

9.4. L_1 távolság alapú α szintű teszt

Beirlant, Györfi és Lugosi [1994] bizonyították, hogy a

$$\lim_{n \rightarrow \infty} m_n = \infty, \quad \lim_{n \rightarrow \infty} \frac{m_n}{n} = 0,$$

és a

$$\lim_{n \rightarrow \infty} \max_{j=1, \dots, m_n} \mu(A_{nj}) = 0,$$

feltételek esetén

$$\sqrt{n}(L_n - \mathbb{E}\{L_n\})/\sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

ahol $\xrightarrow{\mathcal{D}}$ jelöli a konvergenciát eloszlásban és $\sigma^2 = 1 - 2/\pi$.

Legyen $\alpha \in (0, 1)$. \mathcal{H}_0 esetén

$$\mathbb{P}\{\sqrt{n}(L_n - \mathbb{E}\{L_n\})/\sigma \leq x\} \approx \Phi(x),$$

ezért x küszöbérték esetén a hibavalószínűség

$$\alpha = 1 - \Phi(x).$$

Tekintsük azt a tesztet, amelyik elutasítja \mathcal{H}_0 , ha

$$L_n > \mathbb{E}\{L_n\} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha).$$

Beirlant, Györfi and Lugosi [1994] bizonyították, hogy

$$\mathbb{E}\{L_n\} \leq \sqrt{2/\pi} \sqrt{\frac{m_n}{n}}.$$

Ebből következik a teszt végleges, immáron explicit formája, amikor elutasítjuk \mathcal{H}_0 -t, ha

$$L_n > c_2 \sqrt{\frac{m_n}{n}} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha) \approx c_2 \sqrt{\frac{m_n}{n}},$$

ahol

$$c_2 = \sqrt{2/\pi} \approx 0.798.$$

Akkor ez a teszt aszimptotikusan egy α szintű teszt.

Összehasonlítva a két tesztet, mindegyiknél a kritikus érték arányos $\sqrt{m_n/n}$ -nel úgy, hogy a c_2 és a c_1 együtthatók közül a c_2 a kisebb.

10. fejezet

Homogenitás tesztelése

10.1. A tesztprobléma.

Tekintsünk két \mathbb{R}^d értékű, független $\mathbf{X}_1, \dots, \mathbf{X}_n$ és $\mathbf{X}'_1, \dots, \mathbf{X}'_n$ mintát úgy, hogy mind-egyik minta elemei független, azonos eloszlásúak, és az ismeretlen közös eloszlásokat μ illetve μ' jelöli. Az a nullhipotézis, hogy a két minta homogén, azaz a két eloszlás azonos:

$$\mathcal{H}_0 : \mu = \mu'.$$

Jelölje μ_n illetve μ'_n az empirikus eloszlásfüggvényeket az $\mathbf{X}_1, \dots, \mathbf{X}_n$ illetve $\mathbf{X}'_1, \dots, \mathbf{X}'_n$, minták esetén úgy, hogy

$$\mu_n(A) = \frac{\#\{i : \mathbf{X}_i \in A, i = 1, \dots, n\}}{n}$$

illetve

$$\mu'_n(A) = \frac{\#\{i : \mathbf{X}'_i \in A, i = 1, \dots, n\}}{n}.$$

Az \mathbb{R}^d egy $\mathcal{P}_n = \{A_{n,1}, \dots, A_{n,m_n}\}$ véges partíciója esetén a T_n tesztstatisztika a partíción összehasonlítja a két empirikus eloszlást:

$$T_n = \sum_{j=1}^{m_n} |\mu_n(A_{n,j}) - \mu'_n(A_{n,j})|.$$

10.2. L_1 távolság alapú, erősen konzisztens teszt

A következő tétel a T_n statisztika nagy eltéréseinek a jellemzését adja.

10.2.1. tétel (BIAU, GYÖRFI [2005].) *Tegyük fel, hogy*

$$\lim_{n \rightarrow \infty} m_n = \infty, \quad \lim_{n \rightarrow \infty} \frac{m_n}{n} = 0, \quad (10.1)$$

és

$$\lim_{n \rightarrow \infty} \max_{j=1, \dots, m_n} \mu(A_{nj}) = 0. \quad (10.2)$$

Akkor \mathcal{H}_0 és minden $0 < \varepsilon < 2$ esetén

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\{T_n > \varepsilon\} = -g_T(\varepsilon),$$

ahol

$$g_T(\varepsilon) = (1 + \varepsilon/2) \ln(1 + \varepsilon/2) + (1 - \varepsilon/2) \ln(1 - \varepsilon/2).$$

BIZONYÍTÁS. Csak a

$$\mathbb{P}\{T_n > \varepsilon\} \leq 2^{m_n} e^{-ng_T(\varepsilon)} \leq 2^{m_n} e^{-n\varepsilon^2/4}$$

nem aszimptotikus felső korlátot bizonyítjuk. Tetszőleges $s > 0$ esetén a Markov egyenlőtlenségből következik, hogy

$$\mathbb{P}\{T_n > \varepsilon\} = \mathbb{P}\{e^{snT_n} > e^{sne}\} \leq \frac{\mathbb{E}\{e^{snT_n}\}}{e^{sne}}.$$

A Scheffé miatt

$$T_n = \sum_{A \in \mathcal{P}_n} |\mu_n(A) - \mu'_n(A)| = 2 \max_{A \in \sigma(\mathcal{P}_n)} (\mu_n(A) - \mu'_n(A)),$$

ahol a $\sigma(\mathcal{P}_n)$ halmazosztály a \mathcal{P}_n partíció celláinak összes unióiból áll. Ekkor

$$\begin{aligned} \mathbb{E}\{e^{snT_n}\} &= \mathbb{E}\left\{\max_{A \in \sigma(\mathcal{P}_n)} e^{2sn(\mu_n(A) - \mu'_n(A))}\right\} \\ &\leq \sum_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{2sn(\mu_n(A) - \mu'_n(A))}\} \\ &\leq 2^{m_n} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{2sn(\mu_n(A) - \mu'_n(A))}\} \\ &= 2^{m_n} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{2sn\mu_n(A)}\} \mathbb{E}\{e^{-2sn\mu'_n(A)}\}. \end{aligned}$$

Nyilván

$$\begin{aligned}\mathbb{E}\{e^{2sn\mu_n(A)}\} &= \sum_{k=0}^n e^{2sk} \binom{n}{k} \mu(A)^k (1 - \mu(A))^{n-k} \\ &= (e^{2s}\mu(A) + 1 - \mu(A))^n,\end{aligned}$$

és ehhez hasonlóan \mathcal{H}_0 esetén

$$\begin{aligned}\mathbb{E}\{e^{-2sn\mu'_n(A)}\} &= \sum_{k=0}^n e^{-2sk} \binom{n}{k} \mu(A)^k (1 - \mu(A))^{n-k} \\ &= (e^{-2s}\mu(A) + 1 - \mu(A))^n.\end{aligned}$$

A fentiek miatt

$$\begin{aligned}\mathbb{E}\{e^{snT_n}\} &\leq 2^{m_n} \max_{A \in \sigma(\mathcal{P}_n)} (e^{2s}\mu(A) + 1 - \mu(A))^n (e^{-2s}\mu(A) + 1 - \mu(A))^n \\ &= 2^{m_n} \max_{A \in \sigma(\mathcal{P}_n)} [(e^{2s}\mu(A) + 1 - \mu(A)) (e^{-2s}\mu(A) + 1 - \mu(A))]^n \\ &= 2^{m_n} \max_{A \in \sigma(\mathcal{P}_n)} [1 + \mu(A)(1 - \mu(A))(e^{2s} + e^{-2s} - 2)]^n \\ &\leq 2^{m_n} [1 + (e^{2s} + e^{-2s} - 2)/4]^n \\ &= 2^{m_n} [1/2 + (e^{2s} + e^{-2s})/4]^n.\end{aligned}$$

Ebből következik, hogy

$$\mathbb{P}\{T_n > \epsilon\} \leq \inf_{s>0} \frac{\mathbb{E}\{e^{snT_n}\}}{e^{sn\epsilon}} \leq 2^{m_n} \left[\inf_{s>0} \frac{1/2 + (e^{2s} + e^{-2s})/4}{e^{s\epsilon}} \right]^n$$

Belátható, hogy az infimumot az

$$e^{2s} = \frac{1 + \epsilon/2}{1 - \epsilon/2}$$

választásnál kapjuk, és ekkor

$$\mathbb{P}\{T_n > \epsilon\} \leq 2^{m_n} e^{-ng_T(\epsilon)}.$$

A Pinsker egyenlőtlenségből következik, hogy

$$g_T(\epsilon) \geq \epsilon^2/4$$

tehát

$$\mathbb{P}\{T_n > \epsilon\} \leq 2^{m_n} e^{-n\epsilon^2/4}. \quad (10.3)$$

□

A (10.3) korlátból levezethető egy erősen konzisztens homogenitásteszt:

10.1 következmény (BIAU, GYÖRFI [2005].) *Tekintsük azt a tesztet, amelyik elutasítja \mathcal{H}_0 -t, amennyiben*

$$T_n > c_1 \sqrt{\frac{m_n}{n}},$$

ahol

$$c_1 > 2\sqrt{\ln 2} \approx 1.6651.$$

Tegyük fel, hogy (10.1) teljesül, és

$$\lim_{n \rightarrow \infty} \frac{m_n}{\ln n} = \infty.$$

Akkor \mathcal{H}_0 esetén egy véletlen mintanagyság után 1 valószínűséggel a teszt nem hibázik. A fentiekén túl, ha még a $\mathcal{P}_1, \mathcal{P}_2, \dots$ partíciók sorozata aszimptotikusan finom (lásd (8.5)), akkor \mathcal{H}_1 esetén egy véletlen mintanagyság után 1 valószínűséggel a teszt nem hibázik.

BIZONYÍTÁS. \mathcal{H}_0 esetén (10.3)-ból következik, hogy

$$\begin{aligned} \mathbb{P}\left\{T_n > c_1 \sqrt{\frac{m_n}{n}}\right\} &\leq 2^{m_n} e^{-ng_T(c_1 \sqrt{m_n/n})} \\ &= 2^{m_n} e^{-nc_1^2(m_n/n)/4 + n\mathcal{O}(m_n/n)} \\ &= e^{-(c_1^2/4 - \ln 2 + \mathcal{O}(1))m_n}, \end{aligned}$$

ezért a $m_n/\ln n \rightarrow \infty$ miatt

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{T_n > c_1 \sqrt{\frac{m_n}{n}}\right\} < \infty,$$

és a Borel-Cantelli lemmából következik az állítás első fele. A második felével kapcsolatban ugyanúgy belátható, mint (9.7) esetében, hogy (8.5)-ből következik, hogy

$$\liminf_{n \rightarrow \infty} T_n \geq 2 \sup_B |\mu(B) - \mu'(B)| > 0 \quad (10.4)$$

1 valószínűséggel. □

10.3. L_1 távolság alapú α szintű teszt

A \mathcal{H}_0 esetén a T_n statisztika aszimptotikusan normális:

10.3.2. tétel (BIAU, GYÖRFI [2005].) *Tegyük fel, hogy a (10.1) és a (10.2) feltételek teljesülnek. Akkor \mathcal{H}_0 esetén*

$$\sqrt{n}(T_n - \mathbb{E}\{T_n\})/\sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

ahol $\sigma^2 = 2(1 - 2/\pi)$.

A 10.3.2. tételből levezethető egy aszimptotikusan α szintű teszt:

10.2 következmény (BIAU, GYÖRFI [2005].) *Legyen $\alpha \in (0, 1)$, és $C^* \approx 0.7655$ jelöljön egy univerzális konstans. Tekintsük azt a tesztet, amelyik elutasítja \mathcal{H}_0 , amennyiben*

$$T_n > c_2 \sqrt{\frac{m_n}{n}} + C^* \frac{m_n}{n} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha) \approx c_2 \sqrt{\frac{m_n}{n}},$$

ahol

$$\sigma^2 = 2(1 - 2/\pi) \quad \text{és} \quad c_2 = \frac{2}{\sqrt{\pi}} \approx 1.1284.$$

Akkor a 10.3.2. tétel feltételei esetén a teszt aszimptotikusan α szintű. A fentiekén túl, ha még a $\mathcal{P}_1, \mathcal{P}_2, \dots$ partíciók sorozata aszimptotikusan finom (lásd (8.5)), akkor \mathcal{H}_1 esetén a teszt konzisztens.

BIZONYÍTÁS. A 10.3.2. tétel szerint \mathcal{H}_0 esetén

$$\mathbb{P}\{\sqrt{n}(T_n - \mathbb{E}\{T_n\})/\sigma \leq x\} \approx \Phi(x),$$

ezért x küszöbszint esetén a hibavalószínűség

$$\alpha = 1 - \Phi(x).$$

Ezért az α szintű teszt elutasítja a nullhipotézist, amennyiben

$$T_n > \mathbb{E}\{T_n\} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha).$$

Sajnos $\mathbb{E}\{T_n\}$ függ az ismeretlen eloszlástól, ezért arra egy felső korlátot használunk, és ezzel csökkentjük az elsőfajú hibát. Biau, Györfi [2005] igazolta a következő felső korlátot:

$$\mathbb{E}\{T_n\} \leq c_2 \sqrt{\frac{m_n}{n}} + C^* \frac{m_n}{n},$$

tehát

$$\begin{aligned}\alpha &\approx \mathbf{P} \left\{ T_n > \mathbb{E}\{T_n\} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha) \right\} \\ &\geq \mathbf{P} \left\{ T_n > c_2 \sqrt{\frac{m_n}{n}} + C^* \frac{m_n}{n} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha) \right\}.\end{aligned}$$

Az állítás második fele következik a 10.3.2. tételből.

□

11. fejezet

Függetlenség tesztelése

11.1. A tesztprobléma

Tekintsük $\mathbb{R}^d \times \mathbb{R}^{d'}$ értékű $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ független, azonos eloszlású véletlen vektorpároknak egy sorozatát. Az (\mathbf{X}, \mathbf{Y}) eloszlását jelölje ν , míg μ_1 illetve μ_2 legyen a \mathbf{X} illetve a \mathbf{Y} eloszlása. Azt a hipotézisvizsgálati problémát nézzük, amikor a nullhipotézis szerint \mathbf{X} és \mathbf{Y} függetlenek:

$$\mathcal{H}_0 : \nu = \mu_1 \times \mu_2. \quad (11.1)$$

Jelölje $\nu_n, \mu_{n,1}$ és $\mu_{n,2}$ a $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n), \mathbf{X}_1, \dots, \mathbf{X}_n$ és $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ mintákhoz tartozó empirikus eloszlásokat:

$$\begin{aligned} \nu_n(A \times B) &= n^{-1} \#\{i : (\mathbf{X}_i, \mathbf{Y}_i) \in A \times B, i = 1, \dots, n\}, \\ \mu_{n,1}(A) &= n^{-1} \#\{i : \mathbf{X}_i \in A, i = 1, \dots, n\}, \quad \text{and} \\ \mu_{n,2}(B) &= n^{-1} \#\{i : \mathbf{Y}_i \in B, i = 1, \dots, n\}. \end{aligned}$$

Ha adott az \mathbb{R}^d egy $\mathcal{P}_n = \{A_{n,1}, \dots, A_{n,m_n}\}$ partíciója és az $\mathbb{R}^{d'}$ egy $\mathcal{Q}_n = \{B_{n,1}, \dots, B_{n,m'_n}\}$ véges partíciója, akkor egy L_1 tesztstatisztika segítségével összehasonlítjuk a ν_n és a $\mu_{n,1} \times \mu_{n,2}$ eloszlásokat:

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu_n(A \times B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)|.$$

11.2. L_1 távolság alapú erősen konzisztens teszt

Györfi és van der Meulen [1990] bevezetett egy

$$L_n(\mu_{n,1}, \mu_1) = \sum_{A \in \mathcal{P}_n} |\mu_{n,1}(A) - \mu_1(A)|$$

statisztikát, amelyre Biau és Györfi [2005] bizonyította, hogy minden $0 < \varepsilon$,

$$\mathbb{P}\{L_n(\mu_{n,1}, \mu_1) > \varepsilon\} \leq 2^{m_n} e^{-n\varepsilon^2/2}, \quad (11.2)$$

(lásd 9.2.3. tételt). Most megmutatjuk ennek egy kiterjesztését:

11.2.1. tétel (GRETTON ÉS GYÖRFI [2010].) \mathcal{H}_0 és $0 < \varepsilon_1, 0 < \varepsilon_2, 0 < \varepsilon_3$ esetén

$$\mathbb{P}\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \varepsilon_1 + \varepsilon_2 + \varepsilon_3\} \leq 2^{m_n \cdot m'_n} e^{-n\varepsilon_1^2/2} + 2^{m_n} e^{-n\varepsilon_2^2/2} + 2^{m'_n} e^{-n\varepsilon_3^2/2}.$$

BIZONYÍTÁS. Az $L_n(\nu_n, \mu_{n,1} \times \mu_{n,2})$ -t felülről becsüljük:

$$\begin{aligned} L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) &= \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu_n(A \times B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)| \\ &\leq \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu_n(A \times B) - \nu(A \times B)| \\ &\quad + \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)| \\ &\quad + \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\mu_1(A) \cdot \mu_2(B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)|. \end{aligned}$$

\mathcal{H}_0 esetén

$$\sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)| = 0.$$

Továbbá

$$\begin{aligned}
& \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\mu_1(A) \cdot \mu_2(B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)| \\
\leq & \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\mu_1(A) \cdot \mu_2(B) - \mu_1(A) \cdot \mu_{n,2}(B)| \\
& + \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\mu_1(A) \cdot \mu_{n,2}(B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)| \\
= & \sum_{B \in \mathcal{Q}_n} |\mu_2(B) - \mu_{n,2}(B)| + \sum_{A \in \mathcal{P}_n} |\mu_1(A) - \mu_{n,1}(A)| \\
= & L_n(\mu_{n,1}, \mu_1) + L_n(\mu_{n,2}, \mu_2).
\end{aligned}$$

(11.2) miatt

$$\begin{aligned}
& \mathbb{P}\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \varepsilon_1 + \varepsilon_2 + \varepsilon_3\} \\
\leq & \mathbb{P}\{L_n(\nu_n, \nu) > \varepsilon_1\} + \mathbb{P}\{L_n(\mu_{n,1}, \mu_1) > \varepsilon_2\} + \mathbb{P}\{L_n(\mu_{n,2}, \mu_2) > \varepsilon_3\} \\
\leq & 2^{m_n \cdot m'_n} e^{-n\varepsilon_1^2/2} + 2^{m_n} e^{-n\varepsilon_2^2/2} + 2^{m'_n} e^{-n\varepsilon_3^2/2}.
\end{aligned}$$

□

A 11.2.1. tételből levezethetünk egy erősen konzisztens tesztet, amelyik elutasítja a nullhipotézist, ha $L_n(\nu_n, \mu_{n,1} \times \mu_{n,2})$ nagy.

11.1 következmény (GRETTON, GYÖRFI [2010].) *Tekintsük azt a tesztet, amelyik elutasítja a \mathcal{H}_0 -t, amennyiben*

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > c_1 \left(\sqrt{\frac{m_n m'_n}{n}} + \sqrt{\frac{m_n}{n}} + \sqrt{\frac{m'_n}{n}} \right) \approx c_1 \sqrt{\frac{m_n m'_n}{n}},$$

ahol

$$c_1 > \sqrt{2 \ln 2} \approx 1.177. \quad (11.3)$$

Tegyük fel, hogy a

$$\lim_{n \rightarrow \infty} \frac{m_n m'_n}{n} = 0, \quad (11.4)$$

és a

$$\lim_{n \rightarrow \infty} \frac{m_n}{\ln n} = \infty, \quad \lim_{n \rightarrow \infty} \frac{m'_n}{\ln n} = \infty, \quad (11.5)$$

feltételek teljesülnek. Akkor \mathcal{H}_0 esetén 1 valószínűséggel egy véletlen mintanagyság után a teszt nem hibázik. Ha

$$\nu \neq \mu_1 \times \mu_2$$

esetén a \mathcal{P}_n és \mathcal{Q}_n partíciók még aszimptotikusan finomak, akkor 1 valószínűséggel egy véletlen mintanagyság után a teszt nem hibázik.

BIZONYÍTÁS. \mathcal{H}_0 esetén a 11.2.1. tételből következik, hogy

$$\begin{aligned} & \mathbb{P} \left\{ L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > c_1 \left(\sqrt{\frac{m_n m'_n}{n}} + \sqrt{\frac{m_n}{n}} + \sqrt{\frac{m'_n}{n}} \right) \right\} \\ & \leq 2^{m_n m'_n} e^{-c_1^2 m_n m'_n / 2} + 2^{m_n} e^{-c_1^2 m_n / 2} + 2^{m'_n} e^{-c_1^2 m'_n / 2} \\ & \leq e^{-(c_1^2 / 2 - \ln 2) m_n m'_n} + e^{-(c_1^2 / 2 - \ln 2) m_n} + e^{-(c_1^2 / 2 - \ln 2) m'_n}. \end{aligned}$$

A (11.5) feltétel miatt

$$\sum_{n=1}^{\infty} \mathbb{P} \left\{ L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > c_1 \left(\sqrt{\frac{m_n m'_n}{n}} + \sqrt{\frac{m_n}{n}} + \sqrt{\frac{m'_n}{n}} \right) \right\} < \infty,$$

és az állítás első fele következik a Borel-Cantelli lemmából. Az állítás második felénél alkalmazzuk a háromszögegyenlőtlenséget:

$$\begin{aligned} L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) & \geq \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)| \\ & \quad - \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu_n(A \times B) - \nu(A \times B)| \\ & \quad - \sum_{B \in \mathcal{Q}_n} |\mu_2(B) - \mu_{n,2}(B)| \\ & \quad - \sum_{A \in \mathcal{P}_n} |\mu_1(A) - \mu_{n,1}(A)|. \end{aligned}$$

A (11.4) feltétel miatt az utolsó három tag tart 0-hoz 1 valószínűséggel, továbbá ugyanúgy belátható, mint (9.7) esetében, hogy (8.5)-ből következik, hogy

$$\sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)| \rightarrow 2 \sup_C |\nu(C) - \mu_1 \times \mu_2(C)| > 0,$$

ahol az utolsó szupremumot az $\mathbb{R}^d \times \mathbb{R}^{d'}$ összes C Borel halmazára vesszük, tehát

$$\liminf_{n \rightarrow \infty} L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) \geq 2 \sup_C |\nu(C) - \mu_1 \times \mu_2(C)| > 0 \quad (11.6)$$

1 valószínűséggel. □

11.3. L_1 távolság alapú α szintű teszt

Ebben az esetben is van aszimptotikus normalitás:

11.3.2. tétel (GRETTON, GYÖRFI [2010].) *Tegyük fel, hogy a (11.4) és a*

$$\lim_{n \rightarrow \infty} \max_{A \in \mathcal{P}_n} \mu_1(A) = 0, \quad \lim_{n \rightarrow \infty} \max_{B \in \mathcal{Q}_n} \mu_2(B) = 0, \quad (11.7)$$

feltétel teljesül. Akkor \mathcal{H}_0 esetén

$$\sqrt{n} (L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - \mathbb{E}\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2})\}) / \sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

ahol $\sigma^2 = 1 - 2/\pi$.

A 11.3.2. tételből levezethető egy függetlenségi teszt.

11.2 következmény (GRETTON, GYÖRFI [2010].) *Legyen $\alpha \in (0, 1)$. Tekintsük azt a tesztet, amelyik elutasítja \mathcal{H}_0 -t, ha*

$$\begin{aligned} L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) &> c_2 \sqrt{\frac{m_n m'_n}{n}} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha) \\ &\approx c_2 \sqrt{\frac{m_n m'_n}{n}}, \end{aligned}$$

ahol

$$\sigma^2 = 1 - 2/\pi \quad \text{és} \quad c_2 = \sqrt{2/\pi} \approx 0.798.$$

Akkor a 11.3.2. tétel feltételei esetén a teszt aszimptotikusan α szintű. Ha a \mathcal{P}_n és \mathcal{Q}_n partíciók még aszimptotikusan finomak, akkor a teszt konzisztens.

BIZONYÍTÁS. A 11.3.2. tétel szerint \mathcal{H}_0 esetén

$$\mathbb{P}\{\sqrt{n}(L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - \mathbb{E}\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2})\})/\sigma \leq x\} \approx \Phi(x),$$

ezért x küszöbszint esetén

$$\alpha = 1 - \Phi(x).$$

Ekkor egy α szintű teszt elutasítja a nullhipotézist, ha

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \mathbb{E}\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2})\} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha).$$

A $\mathbb{E}\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2})\}$ konstans függ az ismeretlen eloszlástól, de van rá egy felső korlát:

$$\mathbb{E}\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2})\} \leq \sqrt{2/\pi} \sqrt{\frac{m_n m'_n}{n}}$$

(lásd Gretton, Györfi [2010]). □

Irodalomjegyzék

- [1994] Algoet, P. (1994). The strong law of large numbers for sequential decisions under uncertainty. *IEEE Transactions on Information Theory*, 40:609–633.
- [1992] Barron, A. R., Györfi, L., and van der Meulen, E. C. (1992). Distribution estimation consistent in total variation and in two types of information divergence. *IEEE Transactions on Information Theory*, 38:1437–1454.
- [2001] Beirlant, J., Devroye, L., Györfi, L., and Vajda, I. (2001). Large deviations of divergence measures on partitions. *Journal of Statistical Planning and Inference*, 93:1 – 16.
- [1994] Beirlant, J., Györfi, L., and Lugosi, G. (1994). On the asymptotic normality of the l_1 - and l_2 -errors in histogram density estimation. *Canadian Journal of Statistics*, 22:309–318.
- [1946] Bernstein, S. N. (1946). *The Theory of Probabilities*. Gastehizdat Publishing House, Moscow.
- [2010] Biau, G., Bleakley, K., Györfi, L., and Ottucsák, G. (2010). Nonparametric sequential prediction of time series. *Journal of Nonparametric Statistics*, 22:297–317.
- [2005] Biau, G. and Györfi, L. (2005). On the asymptotic properties of a nonparametric l_1 -test statistic of homogeneity. *IEEE Transactions on Information Theory*, 51:3965–3973.
- [1957] Breiman, L. (1957). The individual ergodic theorem of information theory. *Annals of Mathematical Statistics*, 28:809–811.
- [2006] Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press, New York.
- [1952] Chernoff, H. (1952). A measure of asymptotic efficiency of tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–507.
- [1965] Chow, C. K. (1965). Statistical independence and threshold functions. *IEEE Transactions on Computers*, E-14:66–68.
- [1967] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318.
- [1987] Devroye, L. (1987). *A Course in Density Estimation*. Birkhäuser, Boston.

- [1985] Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York.
- [1996] Devroye, L., Györfi, L., and Lugosi, G. (1996). *Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York.
- [2001] Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. Springer-Verlag, New York.
- [2010] Gretton, A. and Györfi, L. (2010). Consistent nonparametric tests of independence. *Journal of Multivariate Analysis*, 11:1391–1423.
- [2002] Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York.
- [2002] Györfi, L. and Lugosi, G. (2002). Strategies for sequential prediction of stationary time series. In *Modeling Uncertainty: An Examination of its Theory, Methods and Applications*, Dror, M., L’Ecuyer, P., and Szidarovszky, F., editors, pages 225–248. Kluwer Academic Publishers, Dordrecht.
- [2007] Györfi, L. and Ottucsák, G. (2007). Sequential prediction of unbounded time series. *IEEE Transactions on Information Theory*, 53:1866–1872.
- [1990] Györfi, L. and van der Meulen, E. C. (1990). A consistent goodness of fit test based on the total variation distance. In *Nonparametric Functional Estimation and Related Topics*, Roussas, G., editor, pages 631–645. Kluwer Academic Publishers, Dordrecht.
- [1963] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30.
- [1969] Kemperman, J. H. B. (1969). On the optimum rate of transmitting information. In *Probability and Information Theory*, pages 126–169. Springer Lecture Notes in Mathematics, Springer-Verlag, Berlin.
- [1999] Kivinen, J. and Warmuth, M. K. (1999). Averaging expert predictions. In *Computational Learning Theory: Proceedings of the Fourth European Conference, Eurocolt’99*, Simon, H. U. and Fischer, P., editors, pages 153–167. Springer.
- [1967] Kullback, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Transactions on Information Theory*, 13:126–127.
- [1933] Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. Roy. Soc. London A*, 231:289–337.
- [1947] Scheffé, H. (1947). A useful convergence theorem for probability distributions. *Annals of Mathematical Statistics*, 18:434–458.
- [1974] Stout, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.